

1-1-1999

Secondary students' skills of measuring liquid volume and understanding of uncertainty of data

Wayne M. Keady
Edith Cowan University

Follow this and additional works at: <https://ro.ecu.edu.au/theses>



Part of the [Science and Mathematics Education Commons](#)

Recommended Citation

Keady, W. M. (1999). *Secondary students' skills of measuring liquid volume and understanding of uncertainty of data*. <https://ro.ecu.edu.au/theses/1209>

This Thesis is posted at Research Online.
<https://ro.ecu.edu.au/theses/1209>

Edith Cowan University

Copyright Warning

You may print or download ONE copy of this document for the purpose of your own research or study.

The University does not authorize you to copy, communicate or otherwise make available electronically to any other person any copyright material contained on this site.

You are reminded of the following:

- Copyright owners are entitled to take legal action against persons who infringe their copyright.
- A reproduction of material that is protected by copyright may be a copyright infringement. Where the reproduction of such material is done without attribution of authorship, with false attribution of authorship or the authorship is treated in a derogatory manner, this may be a breach of the author's moral rights contained in Part IX of the Copyright Act 1968 (Cth).
- Courts have the power to impose a wide range of civil and criminal sanctions for infringement of copyright, infringement of moral rights and other offences under the Copyright Act 1968 (Cth). Higher penalties may apply, and higher damages may be awarded, for offences and infringements involving the conversion of material into digital or electronic form.

USE OF THESIS

The Use of Thesis statement is not included in this version of the thesis.

**Secondary Students' Skills of Measuring Liquid Volume
and Understandings of Uncertainty of Data.**

Wayne M. Keady B.Ed

A thesis submitted in partial fulfilment of the
requirements for the award of
Master of Education
at the Faculty of Community Services, Education,
and Social Sciences, Edith Cowan University.

February, 1999.

ABSTRACT

This research is a detailed study of students' skills of measuring liquid volume, the decisions they make when planning to collect data and whilst collecting data, and the extent to which they understand the uncertainty associated with the data they collected. These skills and understandings are at the heart of scientific literacy (Duggan & Gott, 1996a). The introduction of the *Working Scientifically* strand in the Australian national curriculum framework and profile of learning outcome statements for science (Australian Education Council, 1994) illustrates the increased emphasis placed in curriculum documents on investigation skills and scientific literacy. The profile of outcome statements describes a progression in these skills and understandings.

This study focused on three groups of three students from each of Years 8, 10 and 12 and their performance on two authentic problem solving investigation tasks. The groups of students were observed performing two different investigation tasks that involved the measurement of liquid volume. Video and audio records were made of the groups' use of equipment and dialogue, observations and debriefing interviews provided data for case studies of the groups and how they conducted the investigations.

The study revealed that the students have poor skills of planning for investigation work, and seemed to lack any form of planning schema. Many students

engaged in no up-front planning and only made planning decisions as they collected their data. Very few of the students conducted replicate trials, and those that did perform replicate trials were unable to give a valid reason for doing so. The skills of measuring liquid volume that were observed, revealed a range of skill levels in all age groups. Many students who cited the correct skills for accurate measurement in debriefing interviews did not demonstrate them whilst conducting the investigation. Students generally displayed a poor understanding of uncertainty. No students averaged results from replicate trials, many did not graph their data, some did not record their data but all were confident of the validity of their conclusions.

There was no observed age-based progression of skill for the measurement of liquid volume, with good and poor technique being observed in all age groups. There was a progression, however, in their understanding of uncertainty. Younger students were extremely confident in their conclusions and were unwilling to concede the effect of error on their data whilst the older students did accept that experimental error would affect their data, but did not concede that this effect was great enough to affect the validity of their conclusions.

DECLARATION

I certify that this thesis does not, to the best of my knowledge and belief:

- (i) incorporate without acknowledgment any material previously submitted for a degree or diploma in any institution of higher education;
- (ii) contain any material previously published or written by another person except where due reference is made in the text; or
- (iii) contain any defamatory material.

Signed

ACKNOWLEDGMENTS

I would like to thank, first and foremost, my supervisor Dr Mark Hackling for his encouragement, patience and professional guidance throughout the period of writing this thesis.

Secondly, I would like to express my gratitude to the students who generously gave up their time to assist in this study.

Finally, I would like to thank my wife, Penny, for her support and encouragement.

TABLE OF CONTENTS

	Page
ABSTRACT	ii
DECLARATION	iv
ACKNOWLEDGMENTS	v
TABLE OF CONTENTS	vi
LIST OF TABLES	ix
LIST OF FIGURES	x
 CHAPTER	
1 INTRODUCTION	1
Background	1
The Problem	3
Rationale and Significance	6
Purpose and Research Questions	9
 2 LITERATURE REVIEW	 10
Scientific Literacy, Measurement Competencies and Understanding of Uncertainty	 10
The Role of Practical work in the Development of Investigation Skills	 11
Investigations in the Science Curricula	14
Development of Expertise	17
Findings from Previous Research	19
Approaches to Assessment of Investigation Skills	22
Case Study as a Method of Research	23

3	METHOD	26
	Design	26
	Participants	26
	Instruments	27
	Procedure	28
	Data Analysis	29
4	RESULTS	31
	Introduction	31
	Overview of Groups' Performances on the Two Investigation Tasks	31
	Summary of Performance	34
	Case Studies	43
	Trends in the Development of Measurement Competencies	73
5	DISCUSSION	80
	Approaches to Planning Investigations	80
	Approaches to Data Collection	90
	Skills of Measurement	97
	Understanding Uncertainty	99
	Developmental Trends	103
6	CONCLUSION	106
	Limitations	106
	Research Findings	107
	Implications for Teaching	112
	Implications for Further Research	114
	REFERENCES	116

APPENDICES

1	Opening Statement to Participants	122
2	Equipment List	123
3	Planning and Report Sheet	124
4	Interview Questions	127

LIST OF TABLES

Number		Page
1	Summary of Groups' Performances on the Nappy Investigation	37
2	Summary of Groups' Performances on the Panadol Investigation	42
3	Results of Measurements Including Discarded Data for the Panadol Investigation: Group 12-Chemistry	72

LIST OF FIGURES

Number	Page
1 Task Statement for the Nappies Problem	27
2 Task Statement for the Panadol Problem	28
3 Results Table for The Nappy Investigation: Group 10-3	47
4 Graph of Results for Nappy Investigation: Group 8-2	53
5 Results Table for Nappy Investigation: Group 12-Physics	57
6 Results Table for Panadol Investigation: Group 8-1	61
7 Graph of Results for Panadol Investigation: Group 8-1	63
8 Table of Results for Panadol Investigation: Group 10-1	67

CHAPTER 1: INTRODUCTION

Background

The 1980s saw a change in emphasis for education in Australia with a push to make learning in schools more relevant to the workplace. This led to a number of reports aimed at identifying priorities in the skills of school leavers and recommendations for areas of priority within education. The Key Competencies Report (Mayer, 1992) outlines seven key competencies necessary for further education and participation in the modern workforce. Six of the seven competencies relate directly to the skills required for successful science investigation work and the remaining competency, *Using Technology*, requires understanding of principles and proficiency in the manipulative skills needed for successful interaction with technology.

The Key Competencies Report (Mayer, 1992) signaled a change in emphasis from content to process outcomes in education. Support for this emphasis on process skills can be seen in the UK (Department of Education and Science, 1983) and Canada (Erickson, Bartley, Meyer & Stavy, 1992) where educators have undertaken to assess and map the attainment of investigation and problem solving skills.

This emphasis on investigation and problem solving skills led to the inclusion of the *Working Scientifically* strand in the Australian national curriculum framework and profile of learning outcome statements for Science (Australian Education Council, 1994)

and a similar process strand in the seven other learning areas of the national curriculum. Whilst the strands of outcome statements imply a linear progression in the attainment of understandings and investigation skills there is much debate as to the validity of this assumption (Brown, Blondel, Simon & Black, 1995; Duggan & Gott, 1996b; Fensham, 1994).

Duggan and Gott (1996b) argue that science curricula should be targeting scientific literacy for all as well as science for potential scientists. They define scientific literacy as “having a sound knowledge base in major substantive ideas of science and of ideas relating to the collection, validation, representation and interpretation of evidence” (Duggan & Gott, 1996a, p793). Gott, Duggan, Millar and Lubben (1995) suggest that an informed public needs to be able to enter the debate and evaluate evidence, particularly when judgments are made in matters affecting their lives. This emphasis on scientific literacy has been reflected in the Western Australian Curriculum Framework (Curriculum Council, 1998) in which a large proportion of science outcomes are devoted to investigation skills, communicating scientific understanding, applying science in daily life, and understanding the nature of science as a human activity. This emphasis can also be seen in curricula in the USA which has set a goal of all students achieving scientific literacy (National Academy of Sciences and National Research Council, 1996). It has also been argued that curricula for the new millennium should give priority to problem solving ability, personal effectiveness, ability to communicate and to use technology, and being numerate and scientifically literate (Science Functional Expert Group of the OECD, 1998).

The National Academy of Sciences and National Research Council (1996) define a scientifically literate person as one who is able to evaluate the quality of scientific information on the basis of its source and the methods used to generate it which implies competency in the skills of critical analysis and an understanding of experimental design, techniques of measurement and the uncertainty associated with data (Duggan & Gott, 1996a). The new emphasis on investigation work in the various national curriculum statements is designed to give students opportunities to practise and develop the skills associated with the collection, analysis, interpretation and evaluation of scientific evidence, and to develop understandings of those factors that influence the validity and reliability of data.

The Problem

Despite the greater emphasis on investigation and problem solving skills in Australian and recently developed Western Australian curriculum documents, this push has not yet been translated into working curriculum documents and implemented curricula in Western Australia. The monopoly that content-based outcomes enjoy in secondary science curricula has not, to this point in time, been threatened. Evidence for this can be seen in the syllabuses for Year 11 and 12 Physics in Western Australia. Of 99 statements of outcomes only seven are concerned with any practical work and none refer to specific investigation and problem solving skills (Curriculum Council, 1998).

The Western Australian Monitoring Standards in Education project (Education Department of Western Australia, 1994) revealed that, whilst showing some progression in investigation and problem solving skills through the primary years, secondary students were not progressing in skills development at a satisfactory rate. Within the *Working Scientifically* strand, which has 8 levels and the expectation is that the majority of Year 10 students achieve level 6, students in Year 3 had achieved a mean level of 2-3 and by Year 10 had only progressed to a mean level of 3-4. This modest progression does not, therefore, reflect the emphasis placed on investigation and problem solving skills in curriculum documents despite the relatively high frequency of practical activities in the junior high school curriculum. It seems that the type of laboratory work that is taking place does little to enhance the development of the skills as outlined in the *Working Scientifically* strand.

The aims of laboratory work in schools include the teaching of laboratory skills and to give insight into scientific methods of investigation and develop expertise in using them (Hodson, 1988). However it appears as if many practical activities are missing the mark, and students are not developing the skills in laboratory and investigation work that practical work may set out to develop (Education Department of Western Australia, 1994; Hackling & Garnett, 1995).

The survey of Western Australian lower secondary science teachers undertaken by Staer, Goodrum and Hackling (1998) demonstrated that in over 80% of practical sessions the teacher determined the problem for investigation, provided the equipment

to be used and established the procedural steps to be followed. There is much evidence to suggest that practical work undertaken in this fashion does little to facilitate the acquisition of investigation skills (Hodson, 1988), and this helps explain the poor progression of students' skills in lower secondary science. Hackling and Fairbrother (1996) and Woolnough and Allsop (1985) suggest that the best way to develop investigation and problem solving skills is to give students the opportunity to do open investigation work in science.

A large portion of laboratory work in schools is undertaken to confirm or demonstrate some scientific principle, and as a result teachers feel the need for the demonstration to 'succeed' in order that students gain understanding of the principle (Tamir & Lunetta, 1981). This type of practical work can therefore lead to a large degree of manipulation on the part of the teacher, and as a result does little to develop students' understanding of scientific uncertainty (Fairbrother & Hackling, 1997), which is a vital concept, for the planning and implementation of investigation work. Much of the early decision making in investigation work demands a knowledge of the significance and relevance of scientific uncertainty, and inappropriate decisions made at this stage affect the reliability of collected data and the confidence with which conclusions can be made.

Central to the development of scientific literacy is the acquisition of practical skills and techniques (Hodson, 1988). Considering the importance of measurement skills it is an anomaly to see the low level of attainment of these skills by science students in Western Australia. In a study undertaken to assess the acquisition of investigation skill,

by Year 7 to Year 12 students, it was revealed that nearly half of the students in this age bracket relied on qualitative data when quantitative data were more appropriate (Hackling & Garnett, 1995). Of the students who did take measurements, many made measurements without considering zero values, the range of measurements, parallax error or standardising the measurement procedure (Hackling & Garnett, 1995).

Rationale and Significance

The key skills in investigation work are measuring, manipulating data and interpreting data (Duggan & Gott, 1996a), and these three skills are intrinsically linked to each other. These skills need to be considered together in order that they are meaningful, and they have little value if treated separately (Hodson, 1988). Furthermore these skills of measuring are central to scientific literacy and underpin any discussion of evidence or conclusions drawn from evidence (Hodson, 1988).

The skills associated with the measurement of length, mass, time and liquid volume and a working understanding of the uncertainty that accompanies each type of measurement are essential for investigation work and for living in our technological world. It is these competencies that provide a basis for decision making in the planning phase of an investigation, enable successful data collection to occur and provide the limits for any conclusions drawn from the collected data.

There has been a number of studies of students' acquisition of measurement skills involving length, weight (mass) and force (Brown *et al*, 1993; Duggan & Gott, 1996b; Hackling & Garnett, 1995). However, little research has been conducted into the acquisition of the skills associated with the measurement of liquid volume. Measuring liquid volume is a skill that permeates all disciplines of science at all levels, and is used frequently in the home and in a considerably wide range of occupations, and therefore represents a key skill required for achieving scientific literacy.

Many studies have examined students' understanding of uncertainty in measurement such as Varelas (1997) into fourth grade students' ideas on length measurement based on interviews; Lubben and Millar (1996) into students aged 11, 14 and 16 and their responses to a written survey about the function of repeat measurements; Allie, Buffler, Kaunda, Campbell and Lubben (1998) into the performance of first year university science students' on written probes to do with length measurement; and Sere, Journeaux and Larcher (1993) into the effectiveness of a theoretical course on analysis of measurement errors for first year university students. Whilst these studies probe students' understanding of uncertainty associated with data, only the study by Valeras (1997) was performed within the context of a real life problem; this deficiency of many of the studies was highlighted by Allie *et al* (1998).

Thus there is a gap in the literature in that no study of secondary students undertaking authentic laboratory investigations could be found which examines the

students' understanding of the uncertainty associated with measurement, and certainly none concerning secondary students' skills of measuring liquid volume.

A study of students' acquisition of the skills associated with measuring liquid volume will make a contribution at several levels. The study produced several instruments for the assessment of these skills, and provides useful models for teachers of authentic approaches to assessing measurement competencies in meaningful contexts.

It has been demonstrated through much research that effective teaching requires a knowledge of students' prior knowledge and skills (Osborne, Bell & Gilbert, 1983). This study will provide an insight into the skills that students possess and the level of skill attainment for students at various stages of their schooling. This information will allow teachers to better plan learning experiences aimed at developing skills in investigation work.

The profiles of learning outcomes are in a very early stage of implementation and, as yet, a complete picture of the 'norms' for student achievement has not been established. This study provides some insight into the development of measurement skills and will provide data that can be used to validate some aspects of progression within the *Working Scientifically* strand of the Western Australian student outcome statements (Curriculum Council, 1998).

Purpose and Research Questions

The purpose of this study is to determine how students in Years 8, 10 and 12 approach volume measurements in the context of problem solving laboratory investigations, and if there is any development of skills over these years. The study will focus on the decisions students make about what, how and when to measure, the manipulative skills associated with volume measurement and what consideration students give to measurement uncertainty, and the importance they place upon it, when measuring volume and interpreting their data.

More specifically the study will address the following research questions:

- 1 What decisions are made about the measurement of liquid volume during the planning and data collection phases of an investigation?
- 2 To what extent have students attained competence in measuring liquid volumes, and is there any evidence of progression in skill development from Years 8 to 12?
- 3 To what extent do students understand the uncertainty associated with the volume data they have collected? Does this affect the confidence they have in their conclusions based on these data and is there any evidence of progression in understanding from Years 8 to 12?

CHAPTER 2: LITERATURE REVIEW

Scientific Literacy, Measurement Competencies and Understanding of Uncertainty

The Western Australian Science Learning Area Statement (Curriculum Council, 1998) defines a number of key elements of scientific literacy. They include pursuing initiative and imaginative ideas, responding rationally to events and generating evidence based solutions to problems. The many methods of investigation employed by science are all underpinned by competencies associated with gathering and interpreting scientific evidence (Woolnough & Allsop, 1985). The skills required for successful gathering of information rely on measurement competencies and to successfully interpret gathered information requires an understanding of scientific uncertainty, these competencies are at the core of scientific literacy (Duggan & Gott, 1996a). Scientific literacy is the knowledge and understanding of scientific concepts and processes required for personal decision making, participating in civic and cultural affairs and economic productivity (National Academy of Sciences and National Research Council, 1996). The ability to recognise when fact and opinion are intermingled, or when graphs distort the appearance of results, or when sample size is not reported, or when high levels of precision are unwarranted are key aspects of scientific literacy (Duggan & Gott, 1996a). Tamir and Amir (1987) identified two distinct clusters of skills associated with investigations: analysing problems which involves identifying problems, identifying variables and formulating hypotheses; and, planning experiments which involves deciding on an appropriate experimental design

for a given hypothesis. It is in designing experiments that knowledge of measurement techniques and uncertainty apply, and, as these decisions will determine the validity and reliability of the collected data, these skills are central to investigation work. With a lack of understanding of measurement techniques and the uncertainty associated with them, it is likely that students will have an unwarranted confidence in their conclusions (Hackling & Garnett, 1995).

The Role of Practical Work in the Development of Investigation Skills

To justify the use of practical work in schools it is necessary to examine the aims that can be achieved through practical work. Woolnough and Allsop (1985) identified three fundamental aims that are central to the nature of scientific inquiry and can be validly achieved through practical work: developing practical scientific skills and techniques, developing problem solving skills, and getting a feel for phenomena. Whilst school practical work certainly addresses the development of practical skills and provides an opportunity for students to experience phenomena, there is little opportunity through traditional laboratory work to develop investigation and problem solving skills (Hodson, 1990).

The accepted role of practical work in school science is for motivation and stimulation, teaching laboratory skills, assisting students with learning of scientific principles, developing expertise in the scientific method, and development of scientific attitudes (Hodson, 1988). Garnett, Garnett and Hackling (1995) further

condense the aims of practical work to conceptual learning, techniques and manipulative skills, investigation and problem solving skills, and affective outcomes.

Currently, practical work in W.A. schools is predominantly recipe style activities (Staer *et al*, 1998) conducted with an over-emphasis on conceptual learning, leading to an information overload for the students (Garnett *et al*, 1995). Hodson (1988) warns of the dangers of relying on a single type of learning experience to achieve several different learning aims, and suggests that teachers need to clearly identify the aim of a particular type of practical session so as not to cloud this aim with other less important outcomes. Woolnough and Allsop (1985) identified three different types of practical work and matched these to three aims that can be achieved through practical work; exercises, for developing practical skills and techniques; investigations, for developing problem solving skills; and experiences, to give students a feel for phenomena.

Much of the practical work conducted in schools is concerned with confirming scientific theory or discovering pre-existing ideas and as a result often leads to a 'getting the right answer' mentality amongst students, according to Fairbrother and Hackling (1997). They also suggest that this approach to practical work ultimately leads to students fudging results to fit a preconceived notion of what the results should have been, rather than accepting results and seeking explanation for any discrepancy. Rigano and Ritchie (1995) have reported cases of students fudging results to fit a preconceived 'right' answer. Tamir and Lunetta (1981) argue that to successfully

develop manipulative and investigative skills, laboratory work should be inquiry oriented, with a shift in emphasis from achieving the right answer to achieving solutions to problems. The study by Staer *et al* (1998) indicated that this problem solving type of practical work is not common in Western Australian schools, and is being overlooked for practical work that is very much teacher directed and worksheet based.

Hodson (1988) argues that the teaching of manipulative skills in isolation, removed from any context, has little or no value. He suggests that only skills that are of value in the pursuit of other learning should be taught and that they be taught in the context of the learning to be undertaken (Hodson, 1992). The skills associated with problem solving (problem analysis, planning experiments, collecting, organising and interpreting results) are also best treated in the context of a problem to be solved and hence authentic investigation style laboratory work can provide opportunities for students to learn these skills (Hackling & Garnett, 1995). Hodson (1988) also argues in favour of investigation work to develop understanding of scientific method in addition to traditional 'recipe following' activities and guided discovery style laboratory work. The difference between traditional practical work and investigation work is the extent to which students are involved in making decisions about the processes they are undertaking (Watson & Fairbrother, 1993). Osborne (1993) suggests that the role of practical work should be to provide students with the opportunity to develop skills and to use these skills to undertake genuine investigations, rather than to demonstrate theory.

The Western Australian studies to date confirm that traditional approaches to practical work do little to develop students' investigation skills (Education Department of Western Australia, 1994; Hackling & Garnett, 1995). There is also evidence to support the notion that investigation skills are best developed by doing investigation work. Tamir and Lunetta (1981) found that North American students involved in inquiry style practical work had far more developed problem solving skills than students engaged in more traditional approaches to practical work, they also warn that traditional approaches to practical work can develop a 'warped view of science' amongst students.

The study by Roth and Roychoudhury (1993), conducted in Canada, revealed that higher order process skills are greatly enhanced by inquiry related activities set in authentic contexts and that this enhancement was observed not only in the stronger students but also in weaker students. Tamir and Amir (1987), as a result of research into the inter-relationship between specific skills in biology, warn that knowledge required in order to apply process skills does not necessarily come from manipulating apparatus and materials in the laboratory and that the application of these skills needs to be treated explicitly rather than assumed.

Investigations in the Science Curricula

Staer *et al* (1995) report that teachers in Western Australian schools are reluctant to incorporate investigation work in their classrooms because they believe that students are dependent on the procedural guidance provided by the recipe style

worksheets, they are required to teach so much content that there is insufficient time for investigation work, and they have concerns about managing investigation work in the classroom.

Investigations are activities in which students take the initiative in finding answers to problems (Jones, Simon, Fairbrother, Watson & Black, 1992). The very mechanism of scientific inquiry is inherent in the process of investigation work, for without investigations science would not be science (Coles & Gott, 1993). Investigations require students to plan a course of action, collect the necessary data, organise and interpret that data, reach conclusions and communicate those conclusions in an appropriate form (Garnett *et al*, 1995).

Duggan and Gott (1996a) cite evidence that the rate of progression of children's abilities to conduct investigations appears to decline in the early years of secondary school, and they argue that science education needs to remedy this situation. They suggest that improving the following skills is the key to improving students performance: the ability to generate their own ideas and hypotheses, design an investigation, conduct the data collection, and evaluate the data (Duggan & Gott, 1996a). Fundamental to these skills is an understanding of the uncertainty associated with data that is collected. An understanding of uncertainty will guide the planning of data collection and ultimately determine the validity of conclusions (Woolnough & Allsop, 1985).

Fairbrother and Hackling (1997) detail specific skills and understanding related to uncertainty that should be developed through investigation work. These include: recognising that there is a degree of uncertainty associated with all data; recognising the sources of uncertainty; strategies for reducing the degree of uncertainty in data; making judgments about discarding data; applying degrees of confidence to conclusions; and limiting conclusions to the conditions under which tests are performed.

The Western Australian Science Learning Area Statement (Curriculum Council, 1998) emphasises that science is a collaborative activity and that teaching should reflect this. Many of the key skills of investigation are best taught in a collaborative environment not only to enhance the skills of investigation but to foster teamwork and supporting the work of others (Coles & Gott, 1993). Students should have an understanding of what science is, what science is not, what science can and cannot do, and how science contributes to culture (National Academy of Sciences and National Research Council, 1996). To this end science investigations should be conducted in authentic contexts and draw on students' own experiences in order to explain and predict events in their daily lives (Curriculum Council, 1998). The ability to do science is more than just the ability to demonstrate competence in the component skills of investigation work (Toh & Woolnough, 1990); rather it is the ability to bring these component skills together and use them effectively to solve real problems in a meaningful setting (Hodson, 1993).

Development of Expertise

Currently much of the teaching of science process skills and scientific method is grounded in the belief that skills are arbitrary and are easily transferable from one context to another; however there is growing evidence that expertise in these skills is very much domain specific and tied to particular contexts (Roth & Roychoudhury, 1993).

The ultimate aim of developing problem solving and investigation skills is to lead students from being novices in the field of scientific inquiry through stages of competence towards expertise. McGaw (1986) suggests that experts and novices differ in three fundamental ways: how they represent the problem, how they resolve the problem and how they control their solution processes.

How an individual represents a problem to be solved is crucial to the overall problem solving process. How the problem is represented will determine the type of data collection to take place and the acceptable degree of uncertainty required. Experts spend much more time on representing the problem and planning than novices and they also tend to focus on key elements of the problem and ignore other irrelevant elements (Chi, Feltovich & Glaser, 1981; Hackling & Garnett, 1995). Experts also tend to represent the problem in terms of abstract principles relevant to the particular discipline involved, whereas novices represent problems in terms of the literal objects involved with the problem (Larkin, 1979; McGaw, 1986). This indicates that expertise is dependent upon rich and deep knowledge of the domain in which the problem is set.

Research has shown that experts often use quite different solution processes to those used by novices (McGaw, 1986). In analysing these processes and determining the most appropriate method for developing expertise Collins, Brown and Newman (1989) have proposed a model of cognitive apprenticeship. Whilst a cognitive apprenticeship approach differs from that of the traditional apprenticeship some key elements remain the same, the most notable being that learning takes place within a functional context. The model emphasises the use of conceptual and procedural knowledge in solving problems situated in the context in which they apply (Collins *et al*, 1989), and teaching becomes a process of modeling and coaching with students being provided with scaffolding to support their decision making. Furthermore the model recognises that students do not readily have access to the cognitive processes of experts through the traditional approaches of observation and mimicry (Collins *et al*, 1989), and therefore must rely on cognitive input from teachers as they develop learner autonomy in the cognitive processes required for solving problems. As competence develops scaffolding is faded to allow the learner to assume autonomy over their problem solving processes.

Garnett (1998) conducted research into the effectiveness of different aspects of the cognitive apprenticeship model of instruction with Year 9 science investigation work. It was found that teacher modelling of investigation strategies was least effective when a whole investigation was modelled, and was more effective when only a few aspects of investigation work were modelled. She also found that scaffolding using structured planning and report sheets was very effective and the

requirement for scaffolding reduced as programs progressed. She also noted that students believed that they learned more about doing investigations from talking with their peers than talking with their teacher which gives further support for conducting investigations using small groups.

Findings from previous research

Research into students' performance in investigations in Western Australia has shown that whilst students demonstrate competency in some skills, they are deficient in others. The Monitoring Standards in Education Project (Education Department of Western Australia, 1994) reported that by the end of the compulsory years of schooling 97% of students had demonstrated the more complex skills of level 3 and mastered the less complex skills of level 4. The concern from these data is that the skills associated with accurate measurement and analysing data with an understanding of the uncertainty involved in measurement, appear at levels 4 and above.

Further to this, research has shown an inadequacy in students' understanding of experimental design and measurement uncertainty. The study by Brown *et al* (1995) of students in Years 2, 4, 6 and 7/8 in the U.K., and their responses to different aspects of measurement, revealed that many students were unable to differentiate between measurement and counting and applied the same degree of accuracy to measured data as that applied to counted data. The study also showed that some students were able to recognise that measurements had a degree of uncertainty

associated with them but were more inclined to explain this as human error rather than as limitations of the measuring equipment.

A key indicator of understanding uncertainty is the recognition of the need to conduct repeat trials and the treatment of repeat trials data. The study by Valeras (1997) into how third and fourth grade students integrated repeat trials into their procedures and subsequently their treatment of the repeat trials data revealed that students had not conceptualised the procedure of repeat trials and finding the best representative of the results. Research in the UK regarding the performance of 15 year old students revealed a general reluctance by students to perform repeat trials, and that the decision to repeat trials or not is related to the context in which measurements are taken (Department of Education and Science, 1985). Allie *et al* (1998) showed that undergraduate physics students had difficulty explaining their use of repeat trials, and in many instances were unable to distinguish between systematic and random error.

A key skill in collecting and interpreting evidence is the ability to conduct fair tests which implies an understanding of the importance of isolating the relevant variables and controlling others (Duggan & Gott, 1996a). Varelas (1997) reported that third and fourth grade students had difficulty in explaining fluctuations in data in terms of systematic versus random errors and were haphazard in their use of the terminology they used to describe these errors. This was also evident in the study by Sere *et al* (1993) who reported that undergraduate physics students differentiated poorly between systematic and random error. Even students that have apparently

controlled variables in their investigations may be operating under a 'keep everything the same' mentality without being able to identify controls which have been applied (Department of Education and Science, 1985). Lucas and Tobin (1987) point out that adopting an approach of 'keep everything the same' to the control of variables does not help students to make decisions about controlling relevant variables, or to recognise the possibility of interaction between variables. Allie *et al* (1998) argued that the extent to which students are familiar with the task determines how well skills, such as identifying and controlling variables, are demonstrated. Roth and Roychoudhury (1993) provide support for this point citing that personally meaningful contexts allowed students to develop research skills such as identifying and controlling variables, and that practise within these contexts led to long term retention of those skills.

Hackling and Garnett (1995) showed that secondary students in Western Australia possess poor skills of designing experiments, and their research highlighted a deficiency in students' choices of ranges of measurement and their understanding of measurement error. The study by Duggan and Gott (1996b) examined how secondary students in England and Wales collect and deal with data. Their study revealed a deficiency in students' measurement techniques, in particular appropriate range of measurements, appropriate accuracy of measurements and the use of repeated measurements. A survey of students' measurement skills, from the U.K., highlighted inadequacies in the techniques students used to measure liquid volume, most notably ignoring parallax error, failure to read the bottom of a meniscus, rounding to

convenient labeled graduations and reading scales in the wrong direction (Department of Education and Science, 1985).

Approaches to Assessment of Investigation Skills

Emerging as one of the key issues in investigation style practical work is assessment. The study by Staer *et al* (1995) showed that one of the concerns teachers have about including investigations in their classwork is perceived difficulties in assessing investigation work.

Hodson (1988) warns that assessment of isolated manipulative skills such as measuring and graphing lends credence to the skills being both important in themselves and transferable if the assessment is done free of context. He further suggests that the outcomes of skills such as formulating hypotheses, designing experiments, presenting and interpreting data and drawing conclusions, should be the focus of attention. A method of achieving this is outlined by Hackling and Fairbrother (1996), Watson and Fairbrother (1993) and Toh and Woolnough (1990), whereby students conduct investigations, and record their planning, data collection and data analysis following a prompt sheet. Students record their thinking and doing as they progress through planning and conducting their investigation and analysing their data. Toh and Woolnough (1990) were able to demonstrate a high correlation between students' written reports on these prompt sheets and the records made by an observer. These planning and report sheets scaffold the work of students, reduce the teacher's

management problems and collect data for assessment purposes (Hackling & Fairbrother, 1996).

Learning outcome statements in the *Working Scientifically* strand of the National Profile (Australian Education Council, 1996) and the Western Australian Outcomes and Standards Framework for Science (Education Department of Western Australia, 1998) provide a basis for structuring broadly cued prompt sheets that may be applied to a range of investigation activities and are directly linked with measurable statements. Tamir (1993) warns that assessments that do not match the type of outcome desired are bound to fail and only serve to enhance the 'getting the right answer' mentality, as outlined by Fairbrother and Hackling (1997). Assessment of investigation work should therefore be conducted within the framework provided by profiles of student outcome statements, be used for formative or developmental purposes and conducted within authentic problem solving contexts.

Case Study as a Method of Research

To choose a case study as a method of research is not to choose the method rather it is a choice of object to be studied (Stake, 1994). In contrast to rigid quantitative methods, the case study goes beyond trends in data to establish meaning behind those trends (Stake, 1988). For the field of education, knowledge of how students learn and how schools achieve this learning is the underpinning aim of the educational researcher (Cohen & Manion, 1994), and the simple gathering of statistical outcomes falls short of explaining these aims.

The use of case studies as a research tool has many advantages. Firstly it allows the researcher to adapt to situations that occur and respond to them in a manner that the quantitative researcher cannot, thus an event that may be of significance can be included in the data that would otherwise be excluded from quantitative data (Stake, 1994). Secondly the way students think cannot be accurately represented by responses to survey questions or outcomes on written instruments, it is the processes that are of interest and subjects must be studied in the environment in which these processes occur in order to make the data valid (Stake, 1988). As a consequence the data collected by the case study researcher is much richer and a deeper description of processes that occur than can be provided by quantitative data (Stake, 1994).

The use of case study to probe the thinking of students as they undertake science investigations has been used with success in previous research of this nature, notably Roth and Roychoudhury (1993), Hackling and Garnett (1995), Gott and Duggan (1996), Varelas (1997) and Allie *et al* (1998). The need to probe the complexity of understanding that students have (Varelas, 1997) and the ability to identify 'frames' of approaches to investigation work (Allie *et al*, 1998) are cited as justification for the use of case studies.

Gathering data on the tasks that students perform is a relatively simple process of observation, however gathering data on the decisions students make and the conceptual links behind that decision making is problematic. In order to gather useful data on decision-making processes it is important that students are engaged in a

meaningful problem solving task that is embedded in an authentic context and not just performing tasks that focus on the skills in isolation (Hodson, 1988).

To achieve this participants were given an authentic, problem-solving investigation task to perform, and whilst performing the task in small groups, their dialogue provided insights into their decision making as they worked collaboratively. The data were collected by audio taping the dialogue and a non-participant observer recorded field notes of significant events. Written work samples recorded on a broadly cued planning and report sheet by each group provided evidence about plans for the design of the investigation, data collection, and interpretations made of those data. A debriefing interview, conducted immediately after the investigation was used to probe students' reasons for decisions made regarding experimental design and data collection, and to probe students' understanding of the uncertainty associated with the data they had collected.

This procedure has been used successfully in a number of studies of science investigation processes, including: Duggan and Gott (1996b); Hackling and Garnett (1995); and, Hackling, Garnett, Fairbrother and Tunks (1997).

CHAPTER 3: METHOD

Design

This study consists of a set of case studies of groups of secondary science students involved in a task requiring them to plan and carry out investigations which require measurements of liquid volumes to solve an authentic problem set in a meaningful context. Data were gathered by a non-participant observer collecting field notes as the participants planned and conducted their investigation, audio and video recordings of students' dialogue and apparatus use, and written records made by students. Debriefing interviews were conducted after the students had completed the investigation to allow participants to explain the decisions they made during the task and to probe the students' understanding of the uncertainty associated with the data they collected in their investigation.

Participants

Participants were selected from Years 8, 10 and 12 from an independent coeducational school in a middle class socioeconomic area of the northern metropolitan suburbs of Perth. Case studies were made of three groups of three students from each year level. Intact groups that regularly work together were selected for the study. Exceptionally weak or able students were not included in the sample. Teachers were asked to nominate groups comprising students of average ability, i.e. those achieving grades of C or B. The aim of sampling was to select typical students rather than students from extremes or representative of a particular population. At the

Year 12 level, one group of participants was chosen from students studying Biology, Chemistry and Physics. When possible participants had been selected, parental approval was sought via a letter of informed consent, and any student not gaining approval was replaced in the sample. All of the students approached were successful in gaining approval to participate in the study.

Instruments

Each group was required to complete two investigation tasks involving the measurement of liquid volume.

Task 1: The nappies problem. Participants were given different brands of infant nappies and asked to investigate which of the brands is the most absorbent. To complete this task they were provided with three brands of nappies each of comparable size, a range of graduated beakers, a range of graduated cylinders, a nine litre bucket and an electronic balance. In order that participants were not restricted in their decisions they were instructed to request any additional equipment they required. This task involves the measurement of large volumes of water.

Conduct an investigation to find out which brand of nappy will hold the greatest volume of water/urine. Collect sufficient data that will allow you to confidently advise a friend which brand they should buy as the most absorbent.

Figure 1: Task statement for the nappies problem

Task 2: The Panadol problem. Participants were given a number of soluble pain relief tablets and asked to investigate what effect the volume of water has on the time it takes a tablet to dissolve. To complete this task participants were provided with one packet of soluble pain relief tablets, a range of graduated beakers, a range of graduated cylinders and a stop watch. Again participants were encouraged to ask for any additional items of equipment they needed to complete the task. This task involves the measurement of small volumes of water.

The Panadol company wants to write a recommended volume of water for dissolving soluble Panadol on the side of the box. The company has asked you to investigate the effect of the volume of water on the time taken for the tablet to dissolve. Conduct an investigation so that you may confidently advise the company.

Figure 2: Task statement for the Panadol problem.

Following the introduction to the task statement students engaged in discussion with the researcher as to what the Company required and it was established that they were interested in the volume that would dissolve the tablet most quickly.

Procedure

Prior to commencing the tasks, participants were given the task statement and were invited to ask any questions regarding the tasks, answers were given by the observer to help clarify the tasks but did not indicate procedures that should be followed (see Appendix 1). The groups were supplied with an extensive range of

equipment (see Appendix 2) and instructed to ask for any additional equipment that they needed. All groups completed Task 1 before Task 2. Once the tasks had commenced, with participants working in groups of three, there was no input at all from the observer until the completion of the tasks. Whilst engaged in the tasks participants were required to record their procedures, data and data analysis on a standard planning and report sheet (see Appendix 3) as these sheets formed part of the data analysis for the study.

The groups of students were video and audio recorded whilst performing the tasks, and a non-participant observer recorded field notes of significant events and decisions made by the groups. Each group was interviewed after the second task was completed to allow them to clarify the reasoning behind decisions that they had made whilst carrying out the investigations. The interviews also probed students' understanding of the uncertainty associated with the data they collected in the investigation. The interviews were recorded on audio tape. The interview questions are presented in Appendix 4.

Data Analysis

The data collected via the video and audio recordings were transcribed and catalogued. This was then collated with reference to the field notes and student planning and report sheets, and the groups' performances on several aspects of planning, data collection, data analysis and recommendations, were tabulated for each of the two activities.

The two tables of summary data were analysed separately to identify groups of interest whose performance could be detailed in case studies. The groups were chosen to represent a range of performance for each task. Once these groups were identified the video and audio recordings were scrutinised again in conjunction with the transcripts, field notes and student record sheets in order to construct rich descriptive case studies for individual groups.

The case studies and the summary tables formed the basis for analysing the groups' performances for trends in development of investigation skills. These trends were used to construct developmental continua onto which all of the groups were mapped.

CHAPTER 4: RESULTS

Introduction

The nine groups completed two problem solving tasks each in the style of an investigation and were asked to make recommendations based on the data that they collected. They performed these tasks in groups of three and were video and audio taped while conducting their investigations. As the students worked through each investigation they completed a planning and report sheet. At the completion of the investigation tasks the researcher conducted a debriefing interview with the groups.

This results chapter presents two tables that give an overall summary of the groups' performances on the two tasks. The tables outline the performance of the groups in planning their investigation, the skills that they displayed in measuring liquid volumes while conducting their investigations, and the conclusions and recommendations that they made. Following the tables, three case studies are presented for each task illustrating the range of performance on each task. Finally, from the case studies and in conjunction with the tables, trends were identified and displayed in the form of continua that show progression in performance of competencies associated with the measurement of liquid volume.

Overview of groups' performances on the two investigation tasks

On the nappies investigation, all of the groups adopted the same approach of counting volume lots, that is, pouring measured volumes of water into a nappy as it

either lay on the bench or was held in the hand. This method involved measuring out a number of quantities of equal volume and pouring these onto the nappy. Students recorded the number of pre-measured lots of water that the nappy could hold. Groups varied as to the size of the measured lot and some even changed the size of the measured lot during the trials, but no group attempted to directly measure the volume of water the nappy would hold, by immersing it in a large volume of water. The groups did, however, show variation within the framework of this method to the extent that progression can be observed for several aspects of planning, measurement skills, data collection and data analysis. On occasion, it was not necessarily the case that good performance on one aspect of investigation work, predicted good performance on other aspects. For example, groups that demonstrated planning skills that were more sophisticated than other groups did not necessarily demonstrate good measurement skills or good data analysis skills.

When investigating the effect of volume of water on the time it takes a Panadol tablet to dissolve, all the groups used the same approach to the investigation and variation took place within that general approach, which was to dissolve tablets in different volumes of water and record the time taken for the tablet to dissolve. Many groups pre-determined the volumes that they would use in a search for the volume of water that would give the shortest dissolving time without considering volumes that fall between two measurements or the trend in the data that they collected.

In general terms older participants were more thorough when planning for an investigation and the groups displayed planning behaviour that ranged from no planning at all through to quite extensive and sophisticated planning. A similar trend can be observed in the development of measurement skills where younger students showed no regard for parallax error through to older students who were methodical in their treatment of parallax error.

When groups carried out their data collection there was no discernible difference between the Year 8 and Year 10 groups. For the nappy investigation, groups for both of these Years performed no replication of trials, were inconsistent with the endpoint and did little to control variables. The skills demonstrated by students when conducting the Panadol investigation, the second task, were more sophisticated than those demonstrated on the nappy investigation. However this improvement was evident across the age groups and there was no real progression of skill from Year 8 to Year 12 students in the way in which they collected data.

When making recommendations, younger participants were much more confident in their recommendations than the older participants. Similarly the awareness of the uncertainty associated with the data they had collected increased steadily from the Year 8, through Year 10 to Year 12 students. There was also a strong indication that the younger participants placed more faith in their method of data collection and interpretation of that data than the older participants who were

much more willing to identify shortcomings in their procedures and to concede that measurement and procedural error can affect data.

Summary of performance

The tables that follow provide a summary of the performances of the various groups on the two tasks with reference to planning, measurement skills, data collection, and making recommendations. The categories in the table are explained in more detail on ensuing pages and a legend of terminology is also included. Data reported in Tables 1 and 2 were collated from analysis of videotapes, field notes, written work samples and audio tapes of the debriefing interviews.

The Nappies investigation

Table 1 reveals that some skill areas show improvement across the age groups whilst others do not.

Planning. When planning to take measurements groups did not show any progression of skill across the ages involved and in fact fewer of the Year 12 groups demonstrated any planning for measurement than the younger participants. However it should be noted that when older students did plan to take measurements, the discussion was richer and more detailed than that of younger students. Similarly, the reasons given for the choice of equipment show little improvement across year groups as the reasoning present in younger groups can also be found in the older groups.

Planning of the design and procedure was more sophisticated in the older aged groups, although it was only really in the Year 12 groups that planning for a procedure was treated in any real depth. Whilst some of the younger students hinted at controlling interfering variables it was only the Year 12 groups that approached this aspect methodically, similarly preliminary trials seemed to be considered only by the Year 12 students. The discussions in the planning phase of what constituted a valid end point, that is a point that represented the end of the measurement, seemed confined essentially to the older students and although the younger students did not discuss it during the planning phase some of the groups did discuss this aspect of their procedure as it became an issue during their data collection. None of the groups gave any consideration to sampling error in their planning despite some groups quoting this as a source of error when interviewed.

Skills of measuring volume. The demonstration of measurement skills seems to show no strong developmental trends through the groups as desirable behaviours can be found in both younger and older students and similarly undesirable behaviours may be found in all age groups. Reading a volume by holding a beaker or measuring cylinder up to the level of the eye was not uncommon, similarly resting the container on the bench to ensure that it was vertical and reading the volume from above the container was also prevalent. Although there is some evidence that older students are better able to state that they are reading the bottom of the meniscus, there was no way of checking the video tape to see that this actually occurred. In light of the tendency of some groups to indicate that they had read a volume at eye level or had

the container vertically on the bench, when they clearly hadn't there is some cause to doubt student assurances of reading the bottom of the meniscus and this is more likely to be a learned response.

Collecting data. When collecting data some evidence of progression between ages was observed. The most obvious of these behaviours was the use of replication, which can be found in two of the Year 12 groups but not in any of the younger groups. A tendency towards an awareness of the need for replication was seen in one of the Year 10 groups, as it was suggested by a group member, but the suggestion was not acted on.

Groups that used replicates were also consistent in their end point and made an effort to measure the volume of each aliquot added to the nappy rather than just count volume lots whilst these traits were not observed in groups that did not replicate their tests.

Making recommendations. Some of the clearest progression trends can be seen when the groups were asked to make recommendations. The overall theme in making recommendations was that younger groups were much more confident with their recommendations than the older groups. Subsidiary to this theme was the trend that the younger groups demonstrated very little awareness of errors and the uncertainty associated with their data. It was interesting to note that in two of the Year

Table 1. Summary of groups' performances on the nappy investigation

		Group								
Planning	Measurement	Yr 8 - 1	Yr 8 - 2	Yr 8 - 3	Yr 10 - 1	Yr 10 - 2	Yr 10 - 3	Yr12-phys	Yr12-chem	Yr12-biol
	Reason	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	No
	Control variables	More accurate (beaker)	Easier to pour (mc)	Easier to pour (beaker)	Best way to measure it (mc)	More accurate (mc)	It was just there (beak)	For precision (mc)	Appropriate size (beaker)	Accuracy (mc)
	Pre trials	No	No	Yes	Yes	No	No	Yes	Yes	Yes
	End point	No	No	No	No	No	No	Yes	Yes	No
	Sampling	No	No	No	Yes	No	No	Yes	Yes	Yes
Skills of measuring volume	Bottom of meniscus	No evidence	No evidence	No evidence	Yes	No evidence	No evidence	Yes	Yes	Suggested
	Eye level	No	Said, but not done	Sometimes	Yes	Yes	Yes	Yes	Yes	No
	Vertical on bench	No	Said, but not done	Yes	Yes	No	No	Yes/no	No	Said, but not done
Collecting data	Replication	No	No	No	Suggested	No	No	Yes	Yes	No
	Discarding and repeating	No	No	No	No	No	No	No	No	No
	Consistent endpoint	No	No	No	No	No	No	Yes	Yes	No
	Measurement	No	No	No	No	No	No	Yes	Yes	No
	Care and accuracy	No	No	No	No	No	No	No	No	No
	Sampling							Yes		
	Confirmation								Yes	
Recommendations	Error elimination							Yes	Yes	
	Uncertainty aware	No	No	No	No	No	No	Yes	No	No
	Error aware	No	No	No	Yes	No	Yes	Yes	No	Yes
	Sampling aware	Conflict	Conflict	No	No	Yes	No	Yes	No	Yes
	Based on data	Yes	No data	Yes	Yes	Yes	No data	Yes	Yes	No data

Explanatory notes for Tables 1 and 2

Planning

- **Measurement:** Did groups plan for what equipment they would use to measure volumes of water?
- **Reason:** The reason given for the choice of equipment
- **Control of variables:** Did groups plan for the control of potential interfering variables that may affect the measurements that they would take?
- **Pre trials:** Did groups perform preliminary trials?
- **End point:** Did groups determine an end point for their measurements?
- **Sampling:** Did groups plan for appropriate samples?

Skills

- **Bottom of meniscus:** Did participants demonstrate that they were reading the bottom of the meniscus?
- **Eye level:** Did participants read volumes at eye level?
- **Vertical on bench:** Did participants position the beaker/measuring cylinder vertically by placing it on a flat bench?

Collecting data

- **Replication:** Did groups use replicates/repeat trials?
- **Discarding and repeating:** Did groups discard any initial measurements and repeat the measurements?
- **Consistent end point:** Was the end point for measurements consistent?
- **Measurement:** Did groups take measurements rather than simply counting lots of volumes?
- **Care and accuracy:** Did groups demonstrate care to ensure the accuracy of their measurements?
- **Sampling:** Were replicates/repeat trials performed in consideration of sampling error?
- **Confirmation:** Was replication undertaken to confirm initial data?
- **Elimination of errors:** Was replication undertaken to eliminate errors?

Recommendations

- **Uncertainty aware:** Did groups demonstrate an awareness of uncertainty when making their recommendations?
- **Error aware:** Were groups aware of the affect of errors on their data?
- **Sampling aware:** Did groups demonstrate an awareness of sampling error when making their recommendations?
- **Based on data:** Were recommendations based solely on the data collected?

Legend

Yes: Was demonstrated

No: Was not demonstrated

No evidence: Not possible to ascertain

Conflict: Group could not agree, left undecided

Yes/no: Exhibited by one member of the group but not by other group members

Narrow in: Used a method of gradually refining a volume by working between successively closer extremes

Suggested: Action was suggested by one group member but not acted upon

Said, but not done: An action was not performed but the group said they had done it when interviewed

Sometimes: Action was exhibited by an individual but not consistently

No data: No measurement data were collected, tests were qualitative

Mc: Measuring cylinder

8 groups one of the students suggested the possibility of sampling error but this was rejected by the other members of the group.

The Panadol investigation

Table 2 reveals some similarities and also some marked differences from the information in Table 1.

Planning. Unlike the previous activity, the planning phase for the Panadol investigation did show some progression across the age groups. The reasons given for the choice of the equipment show that older groups tend to be more interested in the accuracy and precision of the equipment whilst the younger groups are more concerned with the convenience of its shape. However, Table 2 reveals that the younger groups tended to plan for the control of interfering variables whilst the Year 12 groups did not, although it was only in the Year 12 groups that any planning of a consistent end point took place in the form of a discussion of what the end point would be and a consensus agreement was reached prior to conducting any trials. The decision about end point was difficult for groups as the tablet rose to the top of the water and broke into several small pieces towards the end of its dissolving time.

Skills of measuring volume. As would be expected the skills of measuring liquid volume demonstrated in this investigation are almost identical to those in the previous investigation. The only exception to this was a group who did not read the meniscus at eye level with the container on the bench in the nappy investigation, did

so on the Panadol task. As with the nappy investigation there were no strong developmental trends in measurement skills across the groups.

Collecting data. When conducting their data collection none of the groups systematically replicated their tests, however some groups disregarded results for some tests and repeated these. There were two reasons groups did this. There were those that started with two extreme volumes and narrowed in to an ideal volume by using progressively closer trial volumes based on the previous lot of trials, and there were those that performed all the trials only once and then identified values that did not fit the general trend and repeated them until they produced a value that did fit. This behaviour was observed in the older groups, Years 10 and 12, but not at all in the Year 8 groups. It was notable that the Year 8 groups tended to display more care to ensure the accuracy of their data than the older groups, however this was coupled with the younger groups taking one trial at a time whilst the older groups were more willing to have several trials being taken concurrently by different group members.

Making recommendations. The progression trends that appear in the nappy investigation for making recommendations can also be seen in the Panadol investigation; however, some intra group discrepancies are evident. Some groups who were prepared to concede sampling error for the nappy investigation did not for the Panadol investigation, whilst some groups indicated an awareness of the uncertainty of the data collected in the Panadol experiment did not indicate this awareness in the

[illegible]

nappy experiment. Overall, the younger groups were much more confident with their recommendations than the older groups as was the case for the nappy investigation.

Case Studies

Selected case studies have been included to illustrate the range of performance observed for the two investigations. Three case studies for each investigation have been chosen, one from each of Years 8, 10 and 12. The case studies were not selected to demonstrate performance progression with age neither have they been chosen to demonstrate typical performance for a particular age group, rather the case study groups illustrate the range of performance and also display some interesting investigation behaviours. The case studies are presented in order of progression of investigation skill and begin with the least developed through to the most developed.

Case Study 1: Nappies investigation, group Year 10 - 3

This group comprised one male and two female students who had worked together prior to their participation in this study. They were chosen as being the top students in a second stream class in a school where there are only two levels of streaming.

The participants were introduced to the task in the standard fashion and commenced their investigation by first examining the different brands of nappies and reading the information provided on the packaging. A discussion of method ensued

but centred mainly on how to tell the different brands of nappies apart during trials “We could try one of each and compare them by the water they take up.” “Are they all the same size?” “No, but we have to compare them.” “But how do we know which one we’re doing...uh hang on they’ve got patterns on them”. It was eventually established that the different patterns would help differentiate them.

“How much do you reckon they will hold?” was a question posed by one participant but subsequently ignored by the others and was not considered again when deciding method. No preliminary trials were conducted to establish the likely range of their measurements. Discussion of equipment was limited to “grab three beakers”, and, although it was not discussed, the three beakers were the same size. It was later revealed that they considered the same size beaker to be more accurate but did not consider the accuracy of measurement afforded by the beaker itself. The decision to use the 200mL beakers, which were marked in 50mL graduations, came from the fact that the graduations on the side were convenient to their method. One of the beakers was filled with water and it was only now that discussion of the method began although this was limited to a suggestion of method that made by one participant and immediately agreed to by the others. The method was to lie each nappy on the bench and pour measured lots of water into them until saturation, which was, as the debriefing interview revealed, based on a method used in television advertisements. The size of the measured lot was discussed after the first beaker was filled with water, and an arbitrary figure of 200mL was decided on, but with no real basis for the figure, “How about 200mL ?” “Sounds good”.

The participant that measured the volumes initially made an attempt to be accurate by holding the beaker at eye level but the beaker was obviously not vertical, and there was no evidence to indicate whether the participant was reading the bottom or top of the meniscus.

When the trial began it was further decided that rather than pour the entire 200mL onto the nappy in one go (aliquot), it would be poured onto the nappy 50mL at a time, with each brand of nappy being tested by one group member. "How much are you pouring in?" "200mL." "All of it?" "Uh, I dunno...make it 50". The participant who had measured the original volumes continued to hold the beaker at eye level but not vertically, whereas the other two participants didn't even hold the beaker at eye level.

After the first lot of 50mL had been added it was suggested and subsequently agreed that the end point (i.e. saturation) would be measured by placing a piece of absorbent paper towel over the nappy and pressing on it after each trial to detect any water that lay unabsorbed on the top of the nappy. "There you go, test it". "How?" "Just feel it." "It doesn't feel wet." "It must be...use the towel and see if any sticks". One group member folded the paper towel into a thickness of four sheets whilst the other two folded theirs into a thickness of two sheets. The nappies were then ranked 1 - 3 (3 being the most absorbent) based on the amount of water appearing on the paper towel. It is important to reiterate that each nappy was being tested by a different group member. It was after the first 50mL had been added, that the group decided to

complete the planning page of the report sheet which, naturally, answered planning questions retrospectively.

After the third lot of 50mL was added the group re-read the task to clarify what it was that they were trying to find out, as they were concerned that their data were not valid. "Are we testing to see how much it absorbs or which one wins?" "To see which one absorbs most." "Should we try another one (50mL aliquot) just to check?" "Yeh, O.K." However on reading the task they were happy that the data they were collecting were sufficient to solve the problem. The tone of the investigation throughout the trial was that of a race; "I think my nappy is going to win" was typical of the approach and this attitude drove both the data collection and the recommendations. There was no record of how much better one nappy was than another, only that it was better based on the lack of water absorbed by the paper towel.

The results recorded by the group consisted of a ranking of the nappies after each measured volume of 50mL had been added. For the first three aliquots the ranking from most absorbent to least absorbent was nappy A, nappy C then nappy B, however for the final aliquot nappy C became the least absorbent, and it was only on this last result that the recommendations were made. The group initially began to construct a graph of their results but when they realised that their data were not conducive to graphing they decided that a graph was not needed to make recommendations.

Record your results here.

	50mL	100mL	150mL	200mL
Nappy A	101	1	1	1
B	13	3	3	2
C	22	2	2	3

Least to most absorbent (1-3)

Figure 3. Results table for the nappy investigation: group 10-3

Before recommendations were made the group made a search of the nappy packaging for any indication of absorbency for the different brands. When no information was found they returned to their data and based their recommendations on the data. There was no analysis of data to identify trends. The group simply decided that nappy A was the “winner” and subsequent answers to questions indicated that nappy A was the winner. There was no real discussion about what recommendations to make and the only discussion centred on how something might be phrased, “The better brands (Huggies) are better to use for babies weighing up to 5kg”.

When discussing the confidence that they had in their results the group were able to identify errors involved with their procedure but maintained that these errors would not have significantly affected their results, and hence had a high degree of confidence in the recommendation that they made. “How important were the errors to

the data that you collected?" "Not very...they wouldn't have changed the nappy one." "Yeh, it was pretty clear cut". Even when questioned in the debriefing interview about repeat trials, the group indicated that if they repeated their trials on additional nappies they would get the same result, as the nappies were all the same, but they gave no indication that they thought repeat trials would highlight or minimise any errors. However the question about repeat trials did prompt one participant to ask "Should we go back and do it again to just check?"

Case Study 2: Nappies investigation, group Year 8-2

The members of this group were selected from a heterogeneous class after their teacher had identified them as being of average ability. The group comprised one female and two male students who had worked as a group together in the past.

After initial instruction the group set to work with one group member writing on the planning and report sheet, whilst the other two examined the available equipment. There was no discussion at all between the two students who checked the equipment but they both appeared purposeful and coordinated in their examination. The silence was at last broken by the writing member asking for help with the planning questions, in particular "What is the plan for your experiment?" Quite an extensive discussion ensued with some jovial and some serious suggestions. "So we just pour the water in there?" "How much water are we putting in?" "Fill it (graduated

cylinder) to about up to 60" "Do we have a prediction?" "No you don't need a prediction". However the discussion took place with consideration of such things as the shape of the nappy, the size of each measured lot of water and the accuracy of the equipment, although their decision on the equipment to use was later revealed to be based on the fact that the measuring cylinder was easier to use and easier to pour from. The group didn't discuss the method to the point of deciding to add measured lots of water to each nappy until saturation, it just seemed to be assumed, but they ultimately made a decision on the method because it was "the easiest and the quickest". After searching the planning and report sheet for a space to put a prediction, and not finding one, they pressed on with their experiment making procedural decisions as they went. These decisions included, holding the nappy over the sink to pour the water in, running the three nappy tests simultaneously, and frequent changes to their determination of what constituted an end point to name a few.

Despite previously discussing the greater accuracy of graduated cylinders compared to beakers, when it came to making volume measurements the cylinders were not at eye level, not vertical on the bench and there was no evidence to determine whether they were reading the bottom or the top of the meniscus. However when asked in the debriefing interview how they ensured their measurements of volume were accurate they cited eye level and vertical on the bench as steps that they had taken. On one occasion the volume was measured by one group member who did put the measuring cylinder on the bench and at least bent down a little to ascertain the

volume, it was then passed to another group member who checked the accuracy of the measurement by holding the cylinder and reading it from above.

The experiment continued with no real consistency of end point with each group member feeling the nappy and commenting on its wetness or dryness, prompting several frank and open discussions on what constituted absorbed and unabsorbed water. "Na that's a bit wet." "But is it going through?" "No but it has to be a good absorber." "But it's still holding it isn't it?" "Yeh, but it's a bit wet". After adding several measured lots, which they changed from 60mL to 70mL, one participant asked "What if they all start dripping at the same time?" The question was ignored by the other group members who appeared to be hoping the situation didn't occur, however it did prompt them to pour their measured volumes onto the nappy more slowly so that if saturation occurred during one lot then the remainder may be accounted for. The group continued to add measured lots to the nappies, still with no consensus as to a consistent end point which appeared to change after each lot was added, then one group member realised that they had lost track of how many measured lots were added to each nappy. They had not been recording the count but the problem was resolved by one group member who was sure that they had added one lot of 60mL and two lots of 70mL to each nappy, with the exception of one nappy that needed 10mL added to make up the correct volume.

The experiment continued with more expressions of uncertainty about how much had been added to a particular nappy and which nappy had received 70mL and

which nappy had received 60mL, but these concerns did not seem to slow the pace of the experiment which bolted toward its ultimate conclusion. The final end point was decided, after supposedly equal amounts of water had been added to each nappy (about 270mL as best as I can determine), by each member of the group feeling the nappy to see which one was the wettest. Once this decision had been made, with no record of results the recommendation was clear and obvious to the point that each group member verbalised their recommendation before they commenced writing the planning stage of the planning and report sheet. "It was the Huggies!" "Huggies won!" "Huggies was the best!"

Once the planning stage was complete the recording of results was decided: stating the "winning" nappy followed by the place getters. The next step was to write down the recommendation "My friend would be told that Huggies 3 layer design is the best absorbent nappy you can buy". This was the prediction that the group had made before data were collected, and while one member of the group wrote this down the other members of the group searched the packaging for supporting evidence. The justification for the recommendations was that they had tested it themselves and were therefore sure of the result, this was reiterated in the debriefing interview when they stated that they were "pretty confident" with their recommendations.

Then came the most difficult task, to construct a graph with no recorded measurements. This had to happen primarily because there was a space for it on the planning and report sheet. Despite the fact that they drew conclusions based on adding

an equal volume of water to each nappy they managed to construct a graph showing different volumes absorbed by each brand of nappy. “How many lots of 60 did you put in?”. “Six”. “No three..... or maybe four”. However they overcame these difficulties to produce a straight line graph of absorbency of different brands of nappies!

In the debriefing interview the participants did recognise sources of error, such as measurement error, procedural error and human error. They readily admitted these errors were important to the data that they had collected, but this did not sway them from their confidence in the recommendation that they had made. However they did recognise that not remembering the number of lots added to the nappies was a possible source of error.

The debriefing interview also revealed a conflict within the group as to whether repeat trials would have yielded the same result. One group member was sure that sampling error would play a part in repeat trials because the nappies would not all be the same whereas the other group members insisted that each nappy of the same brand was exactly the same and therefore the result would be the same. “If you did the experiment again with three nappies of the same type, do you think you would get the same result?” “No, because it depends on the size of the...um...absorber thing.” “The layer?” “Yeh, the cushion thing.” “If you put the same water amount in each nappy then you’re gunna get the same result.” “Might not, ‘cause the layers might be thicker.” “No, but if they’re all the same type of nappy then the layers are gunna be all

the same.” “Yeh, but they can’t be all the same.” “Of three nappies out of the Huggies bag they’re all gunna be the same aren’t they.” “Might not.” “Why?” “Cause...I dunno”. The earlier discussion of error played no part in deliberations as to whether repeat trials would yield the same result or not.

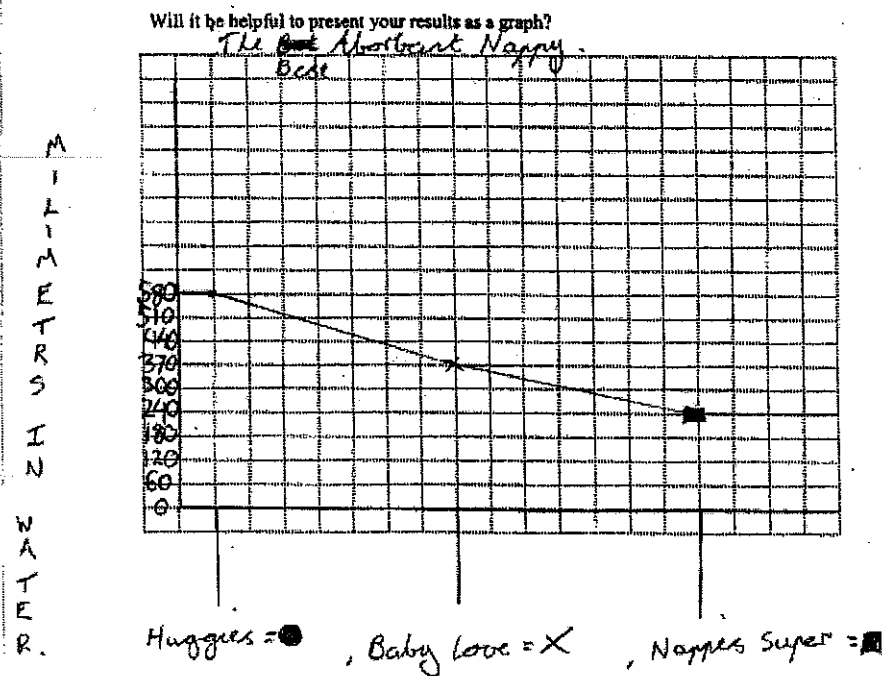


Figure 4. Graph of results for nappy investigation: group 8-2

Case study 3: Nappies investigation, group Year 12 Physics

This group was selected from the only Physics class in the school comprising 12 class members. Their selection was based on their school grade for the subject, a B

or C grade. The group comprised one female and two male students who had worked as a group on class activities.

The group started by examining the three brands of nappies and establishing a method of identifying which nappy was which, as brand names are not written on the nappy itself. They then turned their attention to the selection of equipment from that which was available, examining each piece for its suitability. Then they broke into individual tasks, one group member began writing on the planning sheet, another got a bucket of water whilst the third continued the examination of the equipment.

With the initial examination of the equipment completed, the group now engaged in a discussion of method which focused on such points as what part of the nappy would absorb water, how to add the water to the nappy, whether a preliminary trial was necessary, how the end point of saturation would be determined and how the nappy should be held when adding water to it. They decided on a preliminary trial with one nappy, specifically to determine what the saturation point would look like. The method they decided on was to add measured lots of water to each nappy, although they did discuss other methods but thought that this method was practical. The preliminary trial enabled the group to establish an end point, "Yeh it's rolling off to the sides now." "But is it leaking?" "If it's not leaking it's not finished." "But it's not being absorbed...Oh wait, it's going in now", and it also gave them an indication of volume limits (i.e. the nappy held about 300mL) but despite this they decided to

use lots of 50mL as the other nappies may hold much less than the one used in the preliminary trials.

When measuring the volumes of water two group members placed the measuring cylinders on the bench and read them at eye level the third member held the cylinder in the air at eye level, but all members indicated later that they were reading the bottom of the meniscus. The decision to use measuring cylinders over beakers for volume measurements was taken because they are more reliable and precise, "as close to exact as possible".

Before conducting the experiment the group measured out as many lots of 50mL that the equipment would allow, which meant using up to 250mL measuring cylinders to hold 50mL, and had a bucket of water on standby to refill the measuring cylinders. All the nappies were laid out on the bench and it was decided to complete the saturation of one nappy before starting another. Halfway through the first trial of the first nappy they decided to alter their approach and use smaller volumes of water when they got close to the end point. "Do you think that's all it will take?" "Na, it might take more." "Just pour in 10mL at a time in case it overflows." "That's not how we did the other one." "Well we'll do it like that from here on." No record was kept of individual volume lots added to the nappy but the total amount added was recorded as each nappy trial was completed. As the group suspected that saturation was near they poured a measured volume very slowly so that they may determine the exact amount absorbed by subtracting the remaining volume from the whole lot.

Throughout the first trial there was conflict within the group as to whether the task was to measure the total absorbency of the nappies or the absorbency to the point where the baby would be uncomfortable. "Oh look at that there's a huge hubble there now." " But it's not leaking." "But it's a hubble, it's uncomfortableness for the baby." "It's not asking that." "It's asking which is the most absorbent." "Yeh but look it's through this side bit now." "But it's still not leaking". However science prevailed and total absorbency was decided upon. There was also much discussion as to whether the speed at which the water was poured in would affect the result at all, which brought into question their use of varying volume lots, despite these queries they decided to persist as these factors would be difficult to control. After the first nappy trial the group conducted a very quick and rough repeat trial of that nappy by pouring a litre of water onto the nappy, in the form of four 250mL lots, until the end point had been reached. What ensued was yet another in-depth discussion about the reliability of their method and their ability to control the many variables involved but eventually they seemed to be satisfied that the two results (600mL and 660mL) were similar enough to continue with their original method and the second trial result was discarded.

Although each member of the group was involved in measuring volumes they did make an attempt to control some variables by having each task performed by the same person for each trial and replicate trials. In all, the group replicated each trial twice to have three measurements for each brand of nappy. Throughout the experiment the group made mistakes, like spilling some of the measured lot of water,

and recognised that these would have a bearing on their results but they persisted without discarding any of the trials that were affected by these mistakes.

When the time came to process and present their results they tabulated the results but did not average the figures for the different trials. It was decided that graphing the results was not necessary as the tabulated values gave a clear indication of the most absorbent nappy. "Will it be helpful to present your results as a graph...no the table is obvious." "Maybe we should graph them?" "You don't need a graph to see trends." "We only need to say which one is more absorbent".

Record your results here

	Trial ①	Trial ②	Trial ③
Nappies = 200	580	500	500
Baby love = 600	500	640	600
Huggies =	700	735	690

Will it be helpful to present your results as a graph?

NO TABLE IS OBVIOUS

Figure 5. Results table for nappy investigation: group 12-Physics

When making recommendations about the best nappy the fact that they had done repeat trials made them very confident with their recommendations and stated this in the planning and report sheet, "We conducted 3 trials and Huggies clearly held more". They cited a number of reasons for conducting repeat trials including reduction

of errors, identification of outliers, gaining confidence in their procedure by confirming results, and to account for sampling error. They were prepared to recognise a variety of errors present in the experiment that they had done, citing human error and procedural errors as the most prevalent, in particular the subjective judgment of the end point. They even stated that these errors were vital to the data collected but despite this the data were still clear cut in favor of one particular brand and were still confident in their recommendations.

Case study 4: Panadol investigation, group Year 8-1

The members of this group were selected from a heterogeneous class after their teacher had identified them as being of average ability. The group comprised one female and two male students who had worked as a group together in the past.

The group bypassed the planning and report sheet and went straight to examining the available equipment. The equipment present guided the brief discussion on method as they seemed to have pre-determined a method and were now only considering the size of the measurements that they would take. "O.K. we need...the Panadol." "We got the Panadol." "We'll put it in 250." "Na, this one." "50...40" "Yeh". However they did indicate during the debriefing interview that the choice of method was based on the fact that they couldn't think of another way. There was no examination of the Panadol tablets themselves and certainly no preliminary

trials to establish limits or an end point for the dissolving of the tablet. The volumes of water to be used were solely based on the size of the beakers available, tempered only by a query from one group member as to what constituted a reasonable amount of water to drink, however this was ignored and they included the 900mL volume anyway! The volume measured thus corresponded to the top graduation mark of each beaker (i.e. 900mL, 400mL, 200mL, 80mL, ...etc). There was a suggestion after the experiment that they deliberately chose a logarithmic progression for their volumes but could not say why, "each volume was half of the one before it".

Before a trial had been performed they decided that there were not enough volumes and therefore they needed more. "there's not enough here." "Have we got a 100?" "Yeh, there's one there." "What about 20?" "Uhh...nup." "O.K. we'll do 20". As a result they introduced some graduated cylinders as a method of obtaining different volumes rather than more accurate volumes.

All of the volumes were measured by the same person, who filled the beakers and measuring cylinders directly from the tap. The beakers and cylinders were not put vertical on the bench or read at eye level. There was no evidence that the student was reading the volume from the bottom of the meniscus.

The method they were to embark on was to drop the Panadol tablet into the pre-measured volume of water and measure the time that the tablet took to dissolve using two stopwatches, each to be operated by a different group member. The students

practised using the stopwatches to ensure successful operation of the equipment, but there was no consideration of what would constitute an end point or when timing should start. They discussed whether they would do two volumes at a time and discard the slowest one or whether they would trial each volume individually. "We gotta put the tablets in at the same time." "No we'll do it singularly." "Well if you do them singularly and you put the tablet in..." "Yeh and we'll compare the time." "If you put this one in then that one in they'll be different times...so it will be better if you just do them all at the same time." "No we'll do it singularly". The latter proposal won out as they didn't have enough people to adequately time the dissolving rate if two tablets were dissolving concurrently.

Upon commencing the trials they were very careful to coordinate the dropping of the tablet in the water with the two timers. The decision of end point was taken in turn by group members who indicated by saying "stop" when it was judged that dissolving was complete. They then made the decision to record the result to milliseconds in case the different times were very close and they needed this degree of accuracy to separate them. "2:24?" "Yep" "Do it in milliseconds as well 'cause it could come down to that at the end." "Yeh' so we'll make an average of the milliseconds".

They then continued with each trial with no further discussion apparently in agreement that they had solved all of the problems presented to them. They had started with the largest of the volumes (900mL) and worked their way down. After the

first two trials had been completed they speculated that the smaller the volume the faster the tablet would dissolve. However this trend was not discussed again even when the time to dissolve began to rapidly increase as the volumes got smaller. When they got to the very small volumes (20mL and 10mL in measuring cylinders) they decided that they would not be able to see if the end point had occurred or not so they transferred these volumes into small beakers, but did not re-measure them. On the last volume (10mL) they commented when the time got to four minutes that this would definitely not be the recommended volume; however they persisted with the trial until the tablet had completely dissolved. During the experiment they became a little careless when putting unprotected Panadol tablets onto a wet bench and handling some of the measuring cylinders whilst the tablet was dissolving which had the same effect as stirring.

When analysing the data they decided to draw a graph, but chose to construct a bar graph and treated volume as a discrete rather than a continuous variable.

Record your results here.

900mL	2.38	2.42
400mL	2.17.20	2.17.01
200mL	12.25.27	12.25.10
80mL	12.03.16	12.03.02
40mL	12.48.72	12.48.05
20mL	13.48.53	13.48.72
10mL	15.51.33	15.51.33

Will it be helpful to present your results as a graph?

Dissolving time for panadol...

Figure 6. Results table for Panadol investigation: group 8 - 1.

There was no obvious pattern revealed by the graph. The group decided that their graph showed no trend to base a conclusion on, therefore they chose the volume that dissolved the tablet the fastest and justified this recommendation because it was 14s faster than any other volume that they had used, even though most of the times ranged between 120s and 150s. The tone of voice used when discussing the recommendation was that the best volume was obvious. However, in the debriefing interview they admitted that there may have been a volume between the ones trialed that may have dissolved the Panadol more quickly. Despite this admission they simply reworded their recommendation to say that “out of the ones we tested 80 mL was the best”.

When questioned about sources of error they cited small numbers of trials and accuracy of measurement as errors that may affect their results, but did not concede that the difference would have altered the confidence in their recommendation. “What were the main sources of error in your experiments?” “We should have tested more.” “Could have tested more.” “Yeh.” “And we could have measured more accurate.” “In a graduated cylinder that’s more accurate.” “But it didn’t matter ‘cause the ones that won, won by a fair bit so...” They gave no indication that they were aware of sampling error, but conceded that repeat trials would yield slightly different results due to the errors already cited, but that difference again would not affect the confidence that they had in their recommendation.

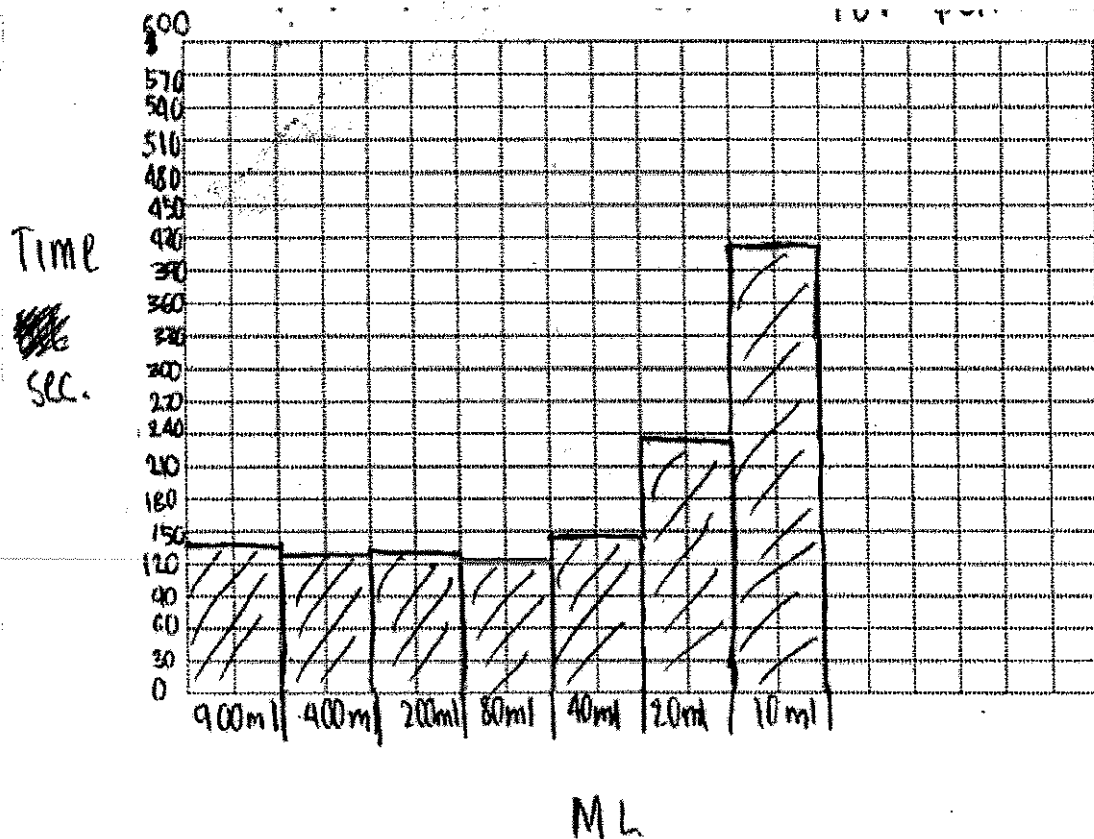


Figure 7. Graph of results for Panadol investigation: group 8 - 1

Case study 5: Panadol investigation, group Year 10-1

This group was chosen from the top streamed class in a system of two levels of streaming. They were chosen as being at the bottom end of the ability range within this class; the group comprised three female students who had worked together in the past.

The group started with a discussion as to whether they should include in their investigation the effect of temperature on the rate of dissolving of Panadol tablets. They eventually decided not to include temperature in their investigation and to treat it as a controlled variable, although one member of the group did not appear convinced even after the decision was made. They then commenced filling out the planning and report sheet whilst simultaneously examining the available equipment. An extensive discussion ensued as to the approach that they should take. "All right Mary (pseudonym) this is your decision, how many of them do we try, and what should we do? Should we just go up in different size beakers, or should we do different amounts or should we use the graduated cylinders?" "Do it in this one...what volume, like what are we gunna use, like how much?" "Well you gotta test it at different amounts." "Yeh I know but what? Like 50, 100, 150...what." "I think...go up in 20s." "In 20s, O.K." "Is there one with 20s?". They decided to start by dissolving a tablet in 20 mL and then increasing the volume by 20mL for subsequent trials. During the discussion one group member suggested that they try one volume to get an idea of the time involved, but the group decided that this was not necessary. "Maybe we should try one just to see how long it will take?" "Yeh but which one, they're all gunna be different." "Just to see where to start." "Na, we just start at the smallest and go up." "There might be one before that." "I don't think so".

The upper limit of their trials was discussed with respect to what constituted a reasonable amount of water to drink but no specific figure was arrived at, they simply agreed to wait and see what trend presented itself as the trials went on, as it was

thought that the relationship might be parabolic producing a convenient minimum value on which they could base their recommendations.

Before commencing the trials one group member measured one volume lot in a beaker (there was no discussion of equipment to use) and placed the beaker vertically on the bench and read it at eye level and they reported in the debriefing interview that they were reading the bottom of the meniscus. Amongst the equipment available they found a pipette and decided to use this to increase the accuracy of their measurements by making small additions or subtractions from each volume with it. "Wait use one of these...then we can just add a little bit" "Oh this, they're cool!" "Oh my God!" "See, now it's perfect". All members of the group used the pipette when adjusting the volumes for trials.

As the trials were about to get underway they discussed whether they should stir the tablet whilst it was dissolving but decided that they would not be able to be consistent in their stirring from trial to trial. "Hang on, are we gunna stir it or anything?" "Umm...no!" "No stirring?" "Because we won't be constant, like each time." "Yeh". The first trial was abandoned and repeated after the person timing had missed the tablet being dropped into the water. While the first trial was underway the third group member completed the planning phase of the planning and report sheet. About 30 seconds into the first trial they nominated one group member to judge the point at which the tablet was fully dissolved, and kept this consistent throughout the experiment. "Yeh, well if we've broken this one in half we need to break every one

we put in there in half.” “Yeh well, we can do that..we shouldn’t have really but oh well.” “It was broken you know, I’m not gunna chuck it away”. The first tablet that they used was broken in the packaging so it was decided that each tablet that they use needed to be broken in half to maintain consistency throughout the trials. It was also during the first trial that they decided to have two trials running concurrently to save time.

This last decision provoked a discussion on whether two different timers might introduce an error but it was eventually agreed that the error involved in timing was negligible and hence the time saving aspect outweighed the small error that might be introduced, even though they cited the stopwatches and timing as a source of error in their experiment. The group made a point of stopping and discarding any trial that was conducted incorrectly, like stopping or starting the stopwatch too late, bumping the beaker and spilling some of the mixture or breaking the tablet into more than two pieces.

When they reached the point of measuring the rate of dissolving in 120mL the group switched to using measuring cylinders with no discussion or explanation as to why. It was about this time that the 100mL trial was completed and the time was the same as for the 60mL trial and both were higher than the 80mL trial, this prompted a large discussion of what was happening and subsequently what to do about it. They suspected that despite the care they had taken they had made mistakes on the 100mL and 60mL trial and that these results should be discarded and the trials repeated. Then

upon reflection they thought that the 80mL trial may have been flawed so they repeated this trial as well. Once the new trials had been done they decided which of the measurements to discard based on how well they fitted the trend apparent in the other data, hence for the 60mL and 80mL trial the second reading was accepted and the first discarded but for the 100mL trial the first reading was retained and the second reading discarded. These decisions were taken as a group and after discussion.

When analysing the data they immediately concluded that they would not be able to justify a recommendation based on the results they had, so they decided to repeat the 60mL and the 100mL trials in an attempt to obtain data that would fit the trend sufficiently so that they may confidently make a recommendation. The new results convinced them to keep the third result for 60mL but retain the original result for the 100mL trial.

Record your results here.

Amount of water	Time in minutes
20 mL	3:0740
40 mL	2:3501
60 mL	2:20 2:33
80 mL	2:12 2:18 2:2474
100 mL	2:2021 2:22
120 mL	2:0554
140 mL	2:135x
160 mL	2:1185

Figure 8. Table of results for Panadol investigation: group 10 – 1

Having decided upon a recommendation based on the results obtained they felt confident enough to write this recommendation on the planning and report sheet. The

final justification for their recommendation was that the trend in the data finally matched the prediction that they had made and was therefore valid, and although they expressed some reservations about their recommendation were happy to stand by it because of the repeats in the trials. When queried about the reason why their results varied for the replicated trials they cited such things as timing error, the shape of the cylinder, in particular the diameter, and accuracy of measuring the time for a Panadol tablet to dissolve.

Case study 6: Panadol investigation, group Year 12 Chemistry

The members of this group were selected from the only Chemistry class comprising 20 students in the school. The students were chosen as they had attained grades of B or C. The group consisted of one male and two female students who had worked together as a group in the past.

The group commenced their investigation with an examination of the equipment presented to them coupled with an initial discussion on method, in particular what volume measurements should they use and how close together they should be. The discussion established that 10mL was far too small and that even 20mL was too small to be practical measurements to use, they then decided that a drinking glass would give them some idea of what limits to set and promptly asked for

the use of one. It was decided that the upper limit of the investigation would be the volume of a drinking glass as it would be impractical to use a greater volume than this under normal circumstances. "We need a cup." "What for?" "So we can measure the maximum volume that the cup would have 'cause they're not going to say two cups!" "Well, what's the average volume of a glass?" "About one cup." "Yeh but.." "About 125mLs." "O.K. then that's the highest we'll go".

They used measuring cylinders to measure the volume of water and also to dissolve the tablet in for the first trial. However, one group member did query this practice but was ignored, and although it wasn't mentioned the first set of trials appeared to constitute preliminary trials for their experiment. They used some chosen volumes (50mL, 70mL and 90mL) as initial tests in the hope that the results from these volumes would give them a clearer idea of what would be the best increment of volume to use for their trials.

When measuring the volumes of water all group members had the measuring cylinders vertical on the bench; however, only two group members made the effort to read the volume at eye level. In the debriefing interview they claimed that they were reading the bottom of the meniscus.

As the trials were about to get underway they decided that instead of measuring the time for a tablet to dissolve in a measuring cylinder they would use a drinking glass but use the measuring cylinders for accurate volume measurement.

“Don’t do it in there ‘cause that will wreck them up, measure in there and pour it in here.” “How much?” “The whole lot, ‘cause that’s more accurate.” “well you wouldn’t want to drink more than that cause it’s gunna taste like dish washing liquid”. The method seemed to be based on their belief that as the volume of water got smaller the time to dissolve got smaller and they continued in this belief until they had gathered several results. They also embarked on a detailed discussion as to how they would make the tests fair and consistent, considering such things as integrity of the tablet, how it was dropped into the water, whether stirring should occur, when timing should begin and a brief discussion on end point, however this discussion yielded no conclusions.

They decided to use a complete tablet and not to stir the mixture, however when the first trials were underway they lifted the glasses and moved them around to try and get a better view of the tablet which had a stirring effect. The first trial established what the end point looked like and they discussed, as it was happening, at what point the tablet was completely dissolved.

The results from the first trial supported their initial prediction that the tablet would dissolve more quickly in smaller volumes although the difference in time was only 10s over a range of about two and a half minutes. They did however decide not to go any lower than 50mL with their measurements because it was an impractical amount of water to drink. “That’s still like chunky.” “Well that was 50.” “Well that’s already saturated that.” “But will people drink it withwell chunks in it?” “Did we

want to go less than 50 'cause mine is definitely less time?" "I think we should do it again 'cause I stuffed mine up....I reckon we go from 50 to 60....see what happens".

The discussion that followed seemed to be more concerned with practical volumes to drink rather than the effect on the tablet's rate of dissolving, thus they spent most of their time trying to establish the volumes they would use. It was suggested that they compare 60mL, 70mL and 80mL but one group member wanted to have a greater difference in volumes to get a better spread of times so they used 60mL, 80mL and 100mL. Despite indicating that their first set of trials were preliminary trials they included the data in their results but then adopted a method of discarding and repeating trials in search of the ideal volume. Their discussion indicated that they were trying to narrow down the ideal volume by using the previous trials as a guide for the next trials. 'Between 50 and 100 there's only like 3 seconds.' "O.K. basically I think we go with the fact that 60mLs dissolves the lot of it with the least time." "Yeh, but so does 90 and 100." "Well we could say between, you never know....shall we do it one more time? Come on and we'll swap so I do the bigger one this time." "I want to do about 200mLs." "That's a whole glass full, no-one's gunna drink a whole glass full." "They will, so let's do that." "O.K. I'll try 150".

The results that they obtained caused them a great degree of concern when trying to make recommendations. The data indicated that as the volume increased so did the dissolving time, although the increase in time from 50 mL to 200 mL was only

16 s. The students failed to consider that for one of the trials the volume of water was topped up with Panadol enriched water. They began to plot the results that they had

Table 3. Results of measurements including discarded data for the Panadol investigation: group 12-chemistry

(a)

1 st attempt	
Volume	Time (min/sec)
50mL	2:12
70mL	2:22
90mL	2:15

(c)

3 rd attempt	
Volume	Time (min/sec)
100mL	2:16
150mL	2:20
200mL	2:28

(b)

2 nd attempt	
Volume	Time (min/sec)
60mL	2:15
80mL	2:22
100mL	2:15

(d)

4 th attempt	
Volume	Time (min/sec)
60mL	2:30
70mL	2:11
80mL	no data

obtained but before it was finished they decided that it wouldn't be any help in making recommendations, and therefore concentrated their efforts on the numbers only. "Nah, it's got nothing to do with volume." "so we're just gunna have to say that now. We haven't exactly...." "The recommended volume over 50, over 50 and then it dissolves, and below 100". They were prepared to accept that there were mistakes made in the first set of trials that made these results unreliable, however they did not delete them from the results table and continued to include them when making conclusions.

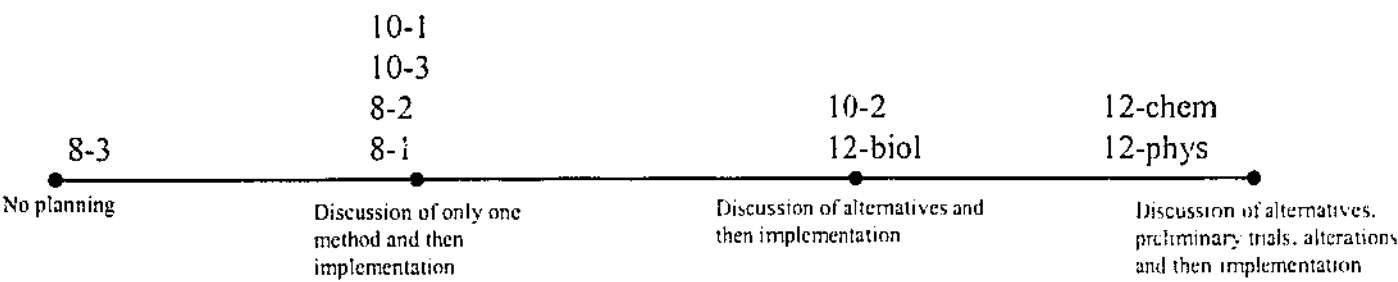
Despite their initial method of trying to narrow down to an ideal volume, they expressed a lack of confidence in choosing a specific volume because of how close the results were. Whilst trying to make a recommendation they discussed reasons for the variation in the results that they had obtained, these included timing error, sampling error and errors associated with their procedure, such as judgment of endpoint, increment of volumes and handling the tablets. Based on these possible errors they recommended a range of volumes as the best (60 - 100mL) rather than identifying a single volume. In the debriefing interview they indicated little confidence in their recommendation and emphasised the importance of timing error to their results. They also indicated that the reason for repeating the measurements that they did repeat (i.e. 60mL and 70mL) was to confirm the original result, and to eliminate personal error by having a different group member time the tablet dissolving.

Trends in the development of measurement competencies

From the analysis of the tables and the case studies it is possible to identify developmental trends for decisions made about planning, skills in making measurements, and understanding of uncertainty in data. These trends have been represented as continua and groups have been located at appropriate positions along the continua. For convenience, the nodes have been located at equal distances apart. The distances between nodes does not represent the extent to which stages of development are different from earlier stages.

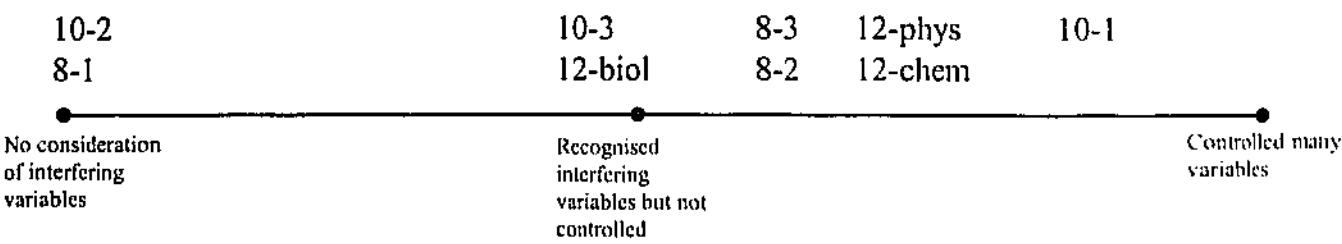
1. Planning the design.

This continuum maps the level of planning that groups engaged in to arrive at their working procedure. The groups are mapped on the basis of their initial planning only; however some groups did alter their procedures once they had commenced data collection when they identified problems.



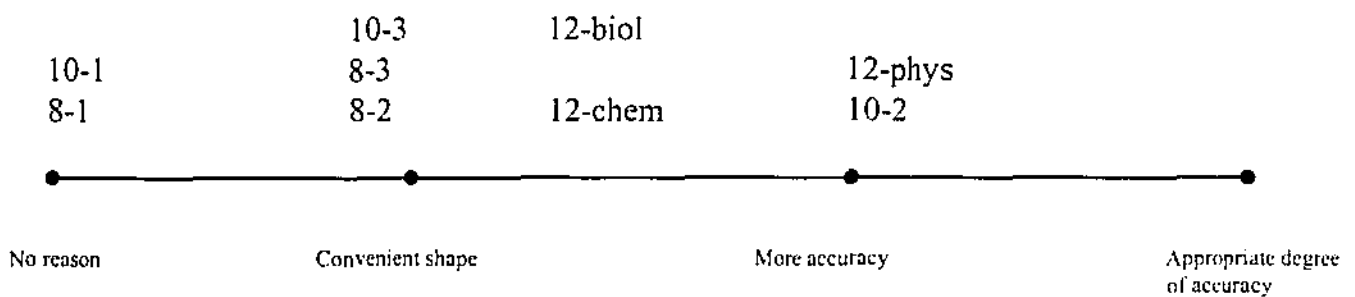
2. Control of variables.

This continuum maps the extent to which groups considered and attempted to control external factors that may affect their data. Some of the groups failed to recognise or control interfering variables whilst other groups identified interfering variables but did not control them. Comments like “that won’t make much difference” or “we can’t do anything about that” were typical of the groups that recognised but did not control interfering variables.



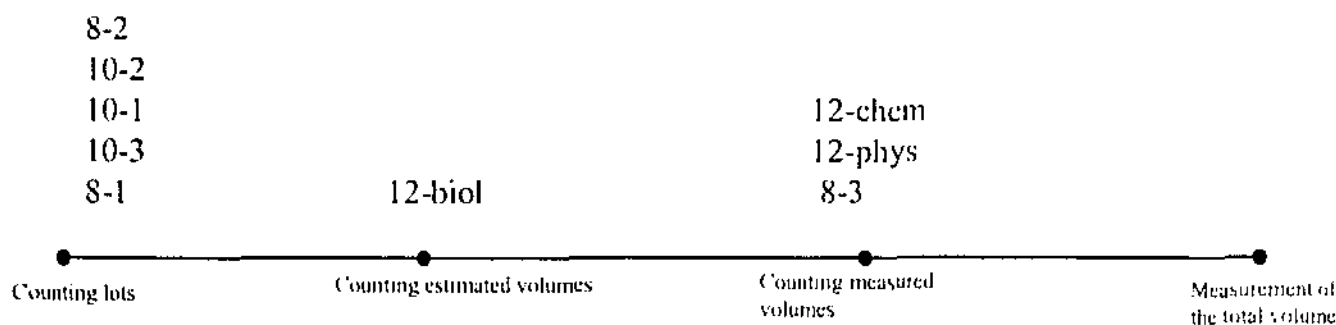
3. Reason for choice of equipment.

This continuum is based on the reasons groups gave for their choice of equipment used to measure volumes when they were asked about it in the debriefing interview. The continuum represents the range of responses given. The final point on the continuum refers to consideration of the degree of accuracy required to measure small or large volumes of water.



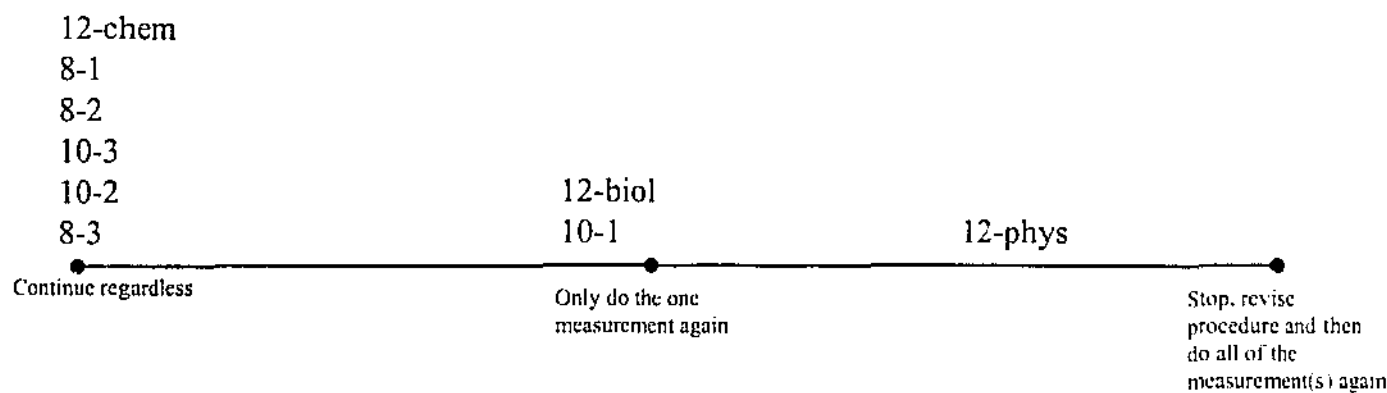
4. Method of measuring the dependent variable in the nappy investigation (volume of water).

This continuum focuses on the way in which groups collected data for the dependent variable, volume of water. Counting refers to a method of using discrete pre-determined volume lots and simply recording the number of those volume lots as their data, whilst measurement refers to determining the actual volume using a measurement scale and thus treating the variable as continuous. There were combinations of the two methods and these are plotted between the extremes on the continuum.



5. Response to mistakes.

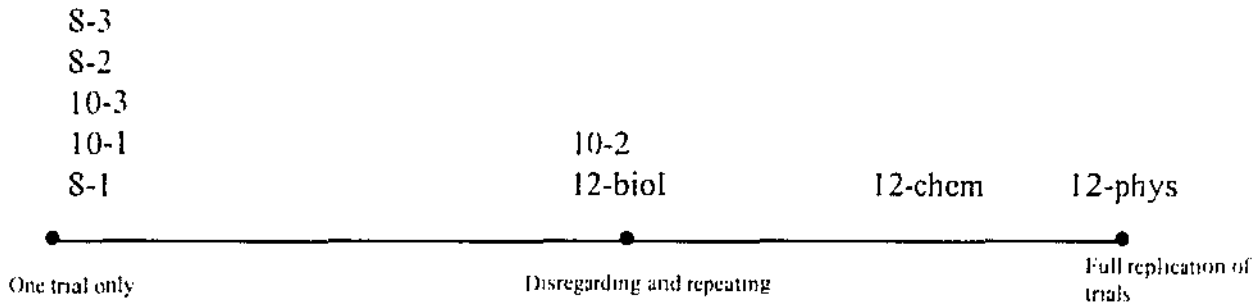
This continuum maps the progression in responses to mistakes made when collecting their data. Mistakes represent gross errors such as spilling some of the water outside the nappy or forgetting to start the stopwatch when the Panadol is introduced to the water.



6. Data collection.

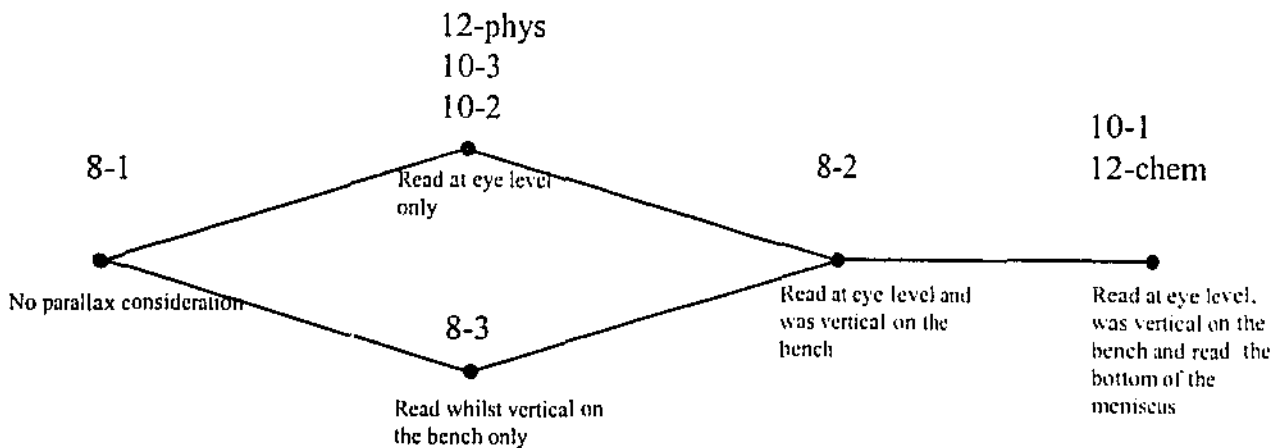
This continuum maps the progression of data collection skills from single trials through to full replication of trials. The middle of the continuum is the stage at which students have gone beyond having reliance on single trials but have not yet recognised the need for complete replication. These students use a method of

identifying data points that do not fit the general trend within the data, disregarding that data point and repeating the trial.



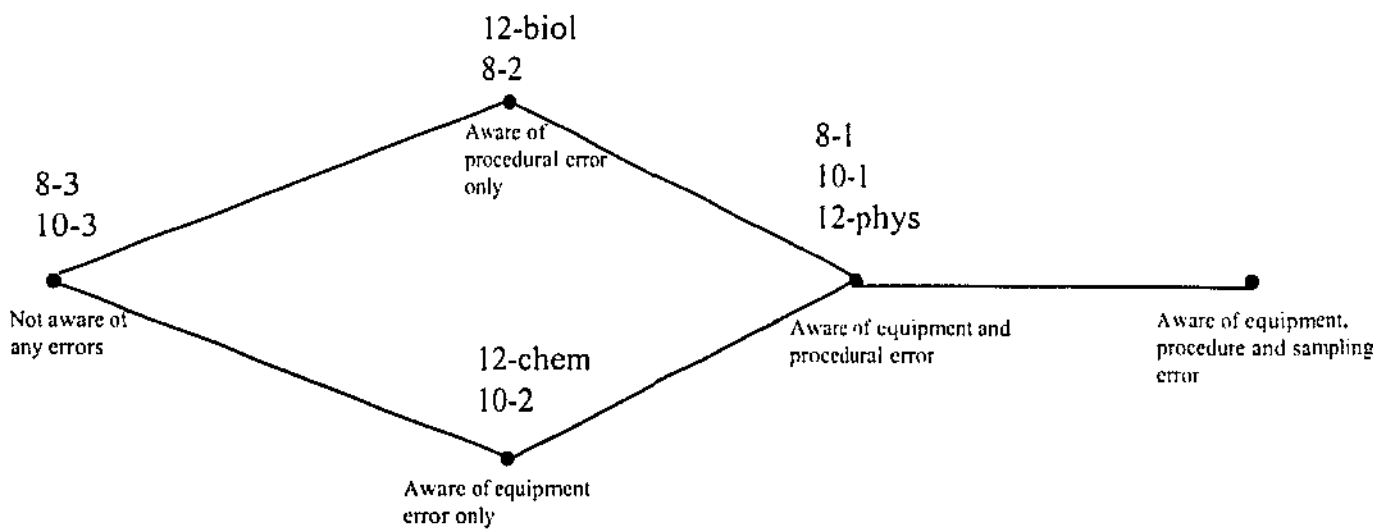
7. Skills of measuring liquid volume.

This continuum represents the progression of skills associated with measuring liquid volume evident within the groups. The middle point of the continuum is not necessarily a linear progression of skills and certainly reading a volume that is vertical on a bench is not dependent on reading at eye level. This is indicated by group 8-3 who placed their container on a vertical bench but did not read it at eye level.



8. Awareness of error associated with data and recommendations.

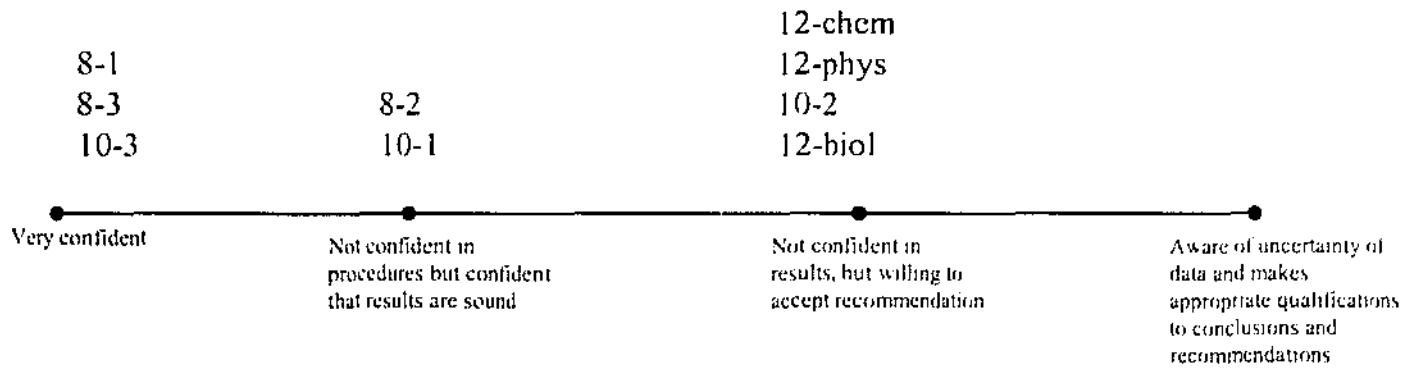
The groups are mapped on this continuum based on the responses they gave during the debriefing interview and what they wrote on their planning and reporting sheet. The continuum maps the extent to which the students were aware of the effect of errors on their data and subsequently their recommendations. Responses indicated that some groups considered the lack of accuracy associated with the equipment to be important to their results, “the beakers aren’t very accurate”, whilst others conceded that their method would have introduced error into the data, “we should have tested more of them”.



9. Confidence in recommendations.

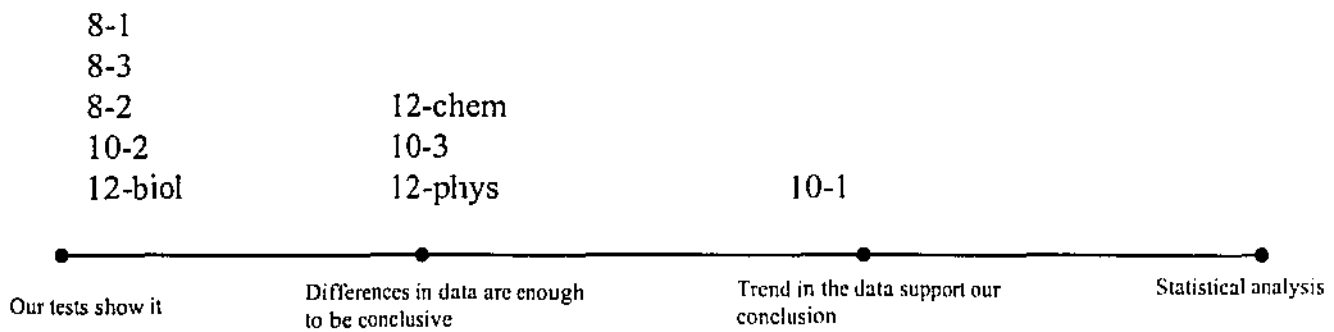
Here the groups are mapped based on the level of confidence in their recommendations that they expressed in the debriefing interview. The continuum is displayed from very confident through to not confident, and this highlights the quite

strong trend of younger groups being more confident of their recommendations than older groups.



10. Reason for confidence.

This continuum was constructed based on the responses given in the debriefing interviews explaining why they were confident in their recommendations and the natural progression that these responses were indicating.



The continua provide a visual representation of the attainment of skills for each of the groups involved and form the basis for the discussion. As such the following chapter should be read with reference to the continua above.

CHAPTER 5: DISCUSSION

The continua provide a basis for discussion in conjunction with the summary tables and the case studies. This chapter will examine the results in five sections. First, the approaches taken by students when planning their investigations will be examined, in particular, how students planned the design for their investigation, their planning of procedure, how they planned to control interfering variables such as sampling error and end point, and their decisions concerning equipment choices. Second, this chapter will examine the approaches taken by students to data collection. Here the focus will be on their responses to mistakes, their method of measuring the independent variable and their use or non-use of replicate trials. Third, this chapter will also examine the skills of measurement demonstrated by the students emphasising measurement of liquid volume. Fourth, the common thread through all of the above sections is the consideration of uncertainty and this chapter will discuss the extent to which students demonstrated an understanding of uncertainty as they conducted their investigations. Fifth, a comparison will be made between the developmental trends represented by the continua and the profiles in the Outcomes and Standards Framework (Education Department of Western Australia, 1998).

Approaches to Planning Investigations

Planning the general approach. The results reveal a range of performance when considering how groups plan their approach, from no planning at all to quite thorough planning. It was common in the younger groups for them to engage in no up

front planning and simply commence data collection, or for one group member to verbalise a method and the others would simply comply with that proposal without any discussion or verbal agreement to do so (Case studies 2 & 4). The planning that younger students did engage in, was generally done during the data collection phase in response to circumstances arising as they worked, similar to that found by Hackling and Garnett (1995). In contrast the older students seemed much more willing and able to present alternative methods and to discuss the merits of each before deciding on an approach.

One of the key aspects to planning investigations is to be able to analyse the problem (Tamir & Amir, 1987), and evidence of this can be seen in some of the groups in this study. In general terms, the Year 8 groups were happy that they had read the problem and that the problem was going to be easy to solve using their pre-determined method. This was typified by one of the Year 8 groups as they undertook the nappy investigation and took measurements of only one trial and did not even record them (Case study 2). The Year 10 groups took a little more care to fully understand that the end result of the task was that a recommendation must be made, however this was not used to guide the planning of their data collection or analysis. The Year 12 groups were better able to let the problem and the need for a recommendation guide their planning of the design and procedure, and their decisions about accuracy and ranges of measurements. This is particularly evident in the Year 12 chemistry group that had asked for a drinking glass in order to ascertain the most

appropriate volume range for the Panadol investigation citing that if a large volume was found to be the best it was impractical to recommend this volume.

The general approach taken to the investigation was remarkably similar for all groups during the nappy investigation and did not vary markedly between groups for the Panadol investigation. It transpired that the choice of method for the nappy investigation was patterned on a television commercial dealing with sanitary pads. This may have occurred as a consequence of the students' lack of exposure to investigation work and as a result they were happy to adopt a common or known method rather than identify their own (Roth & Roychoudhury, 1993). In most cases once the approach had been decided, the investigation as a whole was guided by that approach rather than the investigation being guided by the problem that they were asked to solve.

Planning the method. With respect to planning a method there was no real difference between the Year 8 students and most of the Year 10 students, and the jump in competence only occurred with the Year 12 groups. Continuum 1 demonstrates the development of planning skills present in the sample of students. At one end of the continuum groups did no up-front planning at all or were happy to accept the one method that was proposed, although many groups did make alterations to their design as they proceeded. This result is supported in the findings of Hackling and Garnett (1995) where they showed that high school students tend to identify most

of the interfering variables while they are experimenting. Groups at this level of planning tended to limit their discussions to peripheral aspects of their method and gave little or no consideration to identifying and controlling interfering variables, did not consider the level of accuracy required and did not consider the use of preliminary trials to establish the range and limits of their measurements (Duggan & Gott, 1996a). The lack of awareness by students of the need to consider interfering variables may well be a symptom of the traditional recipe style of practical activity where there is no planning phase for the students, and any consideration of interfering variables is usually dealt with at the end rather than the beginning of the activity (Staer *et al*, 1995). In general, students demonstrated that they lacked a planning schema around which they can conduct investigations and this is evident in the way that students tended to treat each activity separately and evidence of good investigation skills found in one activity for a given group are not evident in the other activity for the same group (Roth & Roychoudhury, 1993).

The other end of the continuum saw groups more inclined to discuss alternatives and even to trial different methods to test the soundness of one method over another. Discussions were centred around the task at hand and attempts were made to control some of the interfering variables. These differences in planning can be highlighted by examining groups control of variables, equipment choices and decisions about end point.

Control of Variables. Similar trends can be seen in the extent to which groups planned to control interfering variables in that older groups were much more willing to spend planning time on identifying variables and planning to control them, whereas the younger groups did little or no planning and tended to deal with them as they became evident during their data collection, if they dealt with them at all (Hackling & Garnett, 1995).

Whilst the younger groups did little or no planning for the Panadol investigation, many of the Year 12 groups were able to discuss, and sometimes control, variables such as subdivision of the tablet, temperature, shape of the container and the effect of agitation. Although these represent a deeper understanding of the interfering variables involved, the higher level consideration of errors tended to come from students studying chemistry, who had previous experience of these issues in their chemistry practical, and not from the students studying biology. It has been suggested that students studying biology may be deficient in some of the skills more commonly associated with chemistry (Tamir & Lunetta, 1981).

Sampling error. One source of error that is unlikely to be dealt with in traditional science laboratory sessions is that of sampling error (Allie *et al*, 1998). None of the nine groups involved, considered this source of error in their planning and even the groups that performed replication of trials made no mention of sampling error in their initial discussions (Groups 12-chem & 12-phys). A common theme throughout the investigations and even the subsequent debriefing interviews was that

students were unwilling to accept that there would be any discrepancy in the manufacture of the nappies or the Panadol tablets. Even the Year 12 chemistry group who had repeated measurements of the same volume and recorded different times were reluctant to consider a difference in the tablets used and were more inclined to offer explanations such as 'human error' or even suggest that volume had no effect on the rate at which a tablet would dissolve.

Planning for Measurement.

When considering the up front planning for taking measurements it was again the older students that considered this aspect of their investigation whilst the younger students tended to make their decisions about measurement as they went along.

Equipment choices. Central to the students' consideration of accuracy is the decisions that they make concerning the equipment that they will use (Education Department of Western Australia, 1994). On Continuum 3 it can be seen that there is a trend towards older students considering the accuracy of the equipment they use whilst the younger students either have no reason for their choice or their reason does not concern accuracy, but the convenience of the shape. Although this trend can be identified in the results it is by no means a strong trend and this is accentuated by the fact that none of the groups, with the possible exception of the Year 12 chemistry group, considered the use of their equipment in terms of the appropriate degree of accuracy for measuring large volumes as opposed to small volumes; rather the only consideration was whether the measuring cylinders were more accurate than the

beakers. The choice of equipment and making decisions about measurements appeared to be performed in isolation without any reference to the whole problem, which Toh and Woolnough (1990) cite as a symptom of skills being taught in isolation free of context.

During planning for their choice of equipment the Year 12 groups gave consideration to the degree of accuracy afforded by the equipment and indeed cited this as the predominant reason for their choice, although some groups had chosen beakers due to their appropriate size or convenient graduations (Groups 12-chem & 12-biol). However only one of the Year 12 groups may have considered the degree of accuracy afforded by equipment, in relation to the recommendation that they were required to make (Group 12-chem). The other Year 12 groups were prepared to cite inaccuracy of equipment as one of their errors, despite having chosen it for the accuracy it afforded. Therefore even though these students could make a link between accuracy and choice of equipment they appear to be of the belief that the more accurate the measurement, the more reliable the data, even if the fluctuations in replicate trials are considerably higher than the tolerance of the equipment being used (Allie *et al*, 1998). These are complex ideas that are often not taught, and the students' lack of ability to reason in terms of appropriate degrees of accuracy is evidence that the traditional practical work that they perform in classes is leading them to search, unnecessarily, for greater degrees of accuracy (Hodson, 1993).

The planning decisions for the Panadol investigation produced some interesting points. Of the groups that actually did plan their measurements, many based the size and range of their measurements on the graduations that appeared on the available equipment rather than being representative of a sensible range with appropriate intervals. This evidence is consistent with Hackling and Garnett's (1995) finding that students in Years 7, 10 and 12 possess poor skills in experiment design and this is highlighted in their poor choices of ranges of measurements.

Further to this, few groups gave consideration to the need to control the shape and size of the measuring containers that they were using and failed to recognise as this as an interfering variable during the Panadol investigation. It was only the Year 12 groups and one of the Year 10 groups that controlled the variable of size and/or shape of the container in which their measurements took place. Two of the Year 10 groups did give consideration to the shape of the container; however this was driven by convenience as the Panadol tablets would not fit into the top of the graduated cylinders that they had initially planned to use (Groups 10-3 & 10-1). Even though these groups had considered the shape of the container they were still prepared to use a range of beaker sizes for their trials.

The overriding issue concerning groups choice of equipment was the extent to which the students considered the degree of accuracy afforded by different types of measuring cylinders. The trend appeared to be that Year 8 groups gave no consideration to the difference in accuracy between measuring cylinders and beakers

and tended to use beakers of varying size randomly or an intermittent combination of beakers and graduated cylinders. Year 12 groups were much more concerned with the issue of accuracy and generally chose their measuring vessel on this basis, with the exception of the Year 12 chemistry group that used beakers to measure volumes for the nappy investigation, however this may indicate that they made a decision concerning the degree of accuracy required for measuring large volumes as the same group used measuring cylinders in the Panadol investigation and cited accuracy as their reason for this choice. The Year 10 groups generally demonstrated an understanding of the difference in accuracy between measuring cylinders and beakers and chose their equipment accordingly, however one Year 10 group (Group 10-3) showed a reluctance to use the measuring cylinders to measure volumes and then pour them into the more convenient beaker for the trials. This was demonstrated when they initially planned to use measuring cylinders for trials in the Panadol investigation and when they realised that the tablet would not fit they reverted to beakers but also took their measurements with the beakers.

Decisions about end point. Perhaps the most critical source of error in the measurements being taken was the consistency of when the measurement had reached its end point. Both in the nappy and panadol investigations the end point of saturation or complete dissolution was not as markedly defined as the students would be used to in their normal practical sessions. Therefore the reliability of the data on which students can base their conclusions is most directly affected by how consistent the end point of measurement has been from trial to trial. This effect can be seen in the range

of data collected from group to group for the same measurement. One of the Year 10 groups measured the Huggies brand of nappy to be saturated with a volume of 800mL (Group 10-1), whilst one of the Year 12 groups were satisfied that the Huggies brand was saturated with 83mL (Group 12-chem). This discrepancy can only be due to the decision, or lack of decision, each group had made about the end point of measurement and the criteria used.

When considering planning for the end point of measurement it is important to note that many groups that had not initially planned their end point did make such decisions after they had commenced their trials when the issue became apparent. However, there is a strong trend in the data that shows that Years 8 and 10 students did not consider end point in their planning, with the exception of one Year 10 group in the nappy investigation, whilst the Year 12 students did plan for their end point before collecting any data. There is also a strong correlation between groups that planned for an end point and groups that conducted preliminary trials, as preliminary trials are a natural progression from the decision to establish an end point.

Despite some groups making an attempt to have consistency to their end point, remarkably few of them were willing to concede this as a source of error in their measurements, indicating a lack of understanding of the uncertainty that this error introduces. This supports the finding by Tamir and Amir (1987) that poor planning may be a consequence of a lack of knowledge of the sources of uncertainty of data.

Approaches to Data Collection.

The review of literature reveals that studies into the competencies students possess in collecting data show deficiencies in controlling variables from trial to trial, measuring the dependent variable (Hackling & Garnett, 1995), identifying and dealing with outlier data points (Brown *et al*, 1995) and awareness of the need to replicate trials (Varelas, 1997). This trend is also revealed in these results with emphasis on three areas, response to mistakes, method of measuring the dependent variable and replication of trials. Another aspect of the data collection phase was the extent to which groups, particularly the younger groups, conducted their planning after they had commenced data collection.

Opportunistic planning. Despite the lack of up front planning undertaken by younger groups many of them did engage in planning as they conducted their investigations. However many decisions made in this manner were more concerned with the practicalities of the procedure than with enhancing the validity of their results

The equipment choices of the younger groups (Years 8 & 10) were predominantly guided by factors other than accuracy or appropriate degrees of accuracy. They chose their equipment on the basis of it having a convenient shape or it was easier to pour from or it had graduation intervals matching those that they intended to use in their data collection (eg. Groups 8-2 & 10-2). It is interesting that consideration of accuracy was not high in priority for these students even for the group that insisted on a timing accuracy of hundredths of a second (Case study 4), and

even the Year 8 group that cited accuracy as a factor in their choice of equipment had chosen the beakers over graduated cylinders (Group 8-1). The difference in accuracy afforded by the scale of graduations between a beaker and a graduated cylinder is stressed in the course that the students undertake during their schooling, however even with this knowledge, it may be that students are reluctant to believe this to be the case, or are oblivious to the possible effect on the data that they collect (Sere *et al*, 1993).

A number of issues emerged from the way groups planned to take their measurements. The Year 8 groups allowed the volumes of the beakers to determine the intervals that they chose to measure as they conducted their investigations, for example, one group chose, for the Panadol investigation, intervals of 20mL, 40mL, 80mL, 200mL, 400mL and 900mL as the intervals for their measurements and these values corresponded to the highest graduation on all of the available beakers (Case study 4).

Some of the on task planning seemed to be done with little consideration to its effect on the actual task. For example, in the Panadol investigation the attempts to control variables, as cited by one Year 8 group, was to ensure that there was a reduction in timing error by having two stop watches timing the one trial (Case study 4). Not only did the students fail to appreciate the level of accuracy required, they expected hundredths of a second to be significant, even though they didn't average the results and graphed the smallest of the times recorded. This demonstrates that the students lack the ability to distinguish between measurement error and interfering

variables as this was the only attempt, that they cited, to control variables for this investigation (Valeras, 1997; Sere *et al*, 1993). It seems as if they were controlling a variable that they had been taught about in class practical activities, without considering the nature of the measurements they were taking in this investigation (Allie *et al*, 1998). This was also evident when the students were questioned in the debriefing interviews about errors, the younger students were inclined to chorus such responses as 'timing error' and 'human error'. Therefore it can be concluded that whilst the students are able to quote their learned errors they do not really understand their effect on the data they are collecting and are certainly not able to make compensation for them when collecting the data. Students need a good understanding of the phenomena they are investigating if they are to identify variables that may affect the outcome of their investigation (Roth & Roychoudhury, 1993).

The decision about what would constitute an end point for their measurements was another aspect that the younger groups considered only after they had commenced data collection. Even after they did establish an end point, the indicators for this point tended to be very subjective. For example, one group tested the absorbency by placing a piece of paper towel on the nappy and made a judgment as to which piece of paper towel appeared least damp (Case study 1). In contrast the older students tended to be more objective in their end point and searched for a point that they felt they could identify with some consistency, for example, when the water began running down the sides of the nappy (Case study 3).

Response to mistakes. Continuum 6 shows that the majority of students who participated in this study, including Year 12 students, were happy to continue with their measurements even though a mistake or procedural error had been identified. Mistakes such as breaking a Panadol tablet in half (Case study 5), forgetting how many aliquots had been added to a nappy (Case study 2), spilling some of a measured volume of water (Case study 3), forgetting to start a stopwatch (Case study 5), placing Panadol tablets onto a wet bench (Case study 4) or using Panadol enriched water (Case study 6) would have affected the data that were collected. Despite this, many groups ignored the effects of these errors and continued with the data collection. This may be a symptom of the traditional approaches to practical work that students undertake in classes where they have a pre-conceived idea of the outcome that they must achieve and are adept at making the data fit that outcome (Rigano & Ritchie, 1995). No better example of this, is the group that did not record the measurements they had taken as to the absorbency of different nappies, yet were able to construct a graph of their non-existent results that clearly confirmed the prediction that they had made prior to collecting any data (Case study 2).

Of the groups that did respond to the mistakes they had made, most of them repeated the measurement they were engaged in when the mistake was identified. The interesting point here is that it was these groups that were much more willing to accept that these mistakes would have influenced their results than the groups that continued regardless. Of the groups that did not repeat trials affected by mistakes, most did concede that this represented an error in measurement but were unwilling to

accept that this would have altered their results enough to change their recommendations.

Method of measuring the dependent variable. The study by Brown *et al* (1995) revealed the difficulty students had in differentiating between measurement and counting and other research has shown that students often reduce a continuous variable to a discrete variable (Duggan, Johnson & Gott, 1996; Garnett, 1998). A similar theme can be identified in these results, particularly in the nappy investigation. Continuum 4 deals with this aspect of the results and shows that no group engaged in measurement of the volume of water absorbed by the nappies and that there were variations in the way in which groups counted lots of water volumes. The most basic of techniques was to add complete lots until saturation occurred (or was perceived to occur) and no partial lots were considered (eg. Case study 1), next was to count lots and to estimate the fraction of one lot remaining (eg. Group 12-biol), through to counting lots and to measure the fraction of one lot remaining (eg. Case study 3). Whilst the Year 12 students exhibited the more sophisticated of the techniques and most of the younger students only counted whole volume lots, one Year 8 group displayed the most sophisticated of the observed techniques.

The method of data collection employed here introduced further uncertainties that tended to be overlooked by most groups. Methods of data collection that were not considered by any of the groups in the nappy investigation included pouring one large volume into the nappy until saturation had occurred and measuring the residual, and

submerging the nappy in a large measured volume and measuring the residual. The inconsistency of measuring the volume lots because of the difference in skill level of the students within a group that were charged with the task, and the compounding error introduced with the increase in the number of measurements being taken were factors not considered by any of the groups. The lack of consideration of these two factors by all of the groups may indicate that the students do not see them as being significant or that they are unable to identify them as a source of error. This may indicate a lack of understanding of the difference between discrete and continuous data (Brown *et al*, 1995). There is some evidence that this may be the case as two of the Year 8 groups and one of the Year 10 groups graphed the results for the Panadol investigation as discrete data instead of continuous data (eg. Case study 4).

Replication of trials. An understanding of the need to replicate trials was only observed in the older students participating in this study. Two of the Year 10 groups and all of the Year 8 groups were happy to base their conclusions and recommendations on single trials.

Only two of the Year 12 groups engaged in full replication of trials. Both Year 12 groups planned to replicate trials but for different reasons, and it was only in the nappy investigation that they replicated trials. When questioned as to why they had replicated trials one group said that they had done it to take into account sampling error and to eliminate or 'even out' errors. "To try and reduce any errors that might have happened in the first one" "Gives an indication of any obvious differences,

outliers and that.." "Because they weren't created perfectly the same" (Group 12-phys). The other group indicated that they had performed replicate trials in order that the extra trials would provide confirmation of the initial data and to eliminate errors. "To get the accurate answer in case something went wrong with the measurements" "To make sure our errors weren't way off with each one" (Group 12-chem). It is interesting to note that neither group chose to apply any statistical treatment to the replicate results, such as averaging, and simply relied on the visual impact of the numbers, which was a subjective judgment about the differences, on which they based their conclusions and recommendations. This is of concern as conducting replicate trials and calculating averages from replicates is listed at level 4 of the *Working Scientifically* strand of the Western Australian Outcomes and Standards Framework (Education Department of Western Australia, 1998). By the end of Year 8 it is expected that a majority of students would have achieved level 4. It is evident that students at Years 8 and 10 are not aware of the need to replicate trials and this may indicate they are not aware of the potential for random errors in these investigations (Allie *et al*, 1998).

A variation on conducting replicate trials was to discard and repeat several, if not all, trials. This method was adopted by all of the Year 12 groups and two of the Year 10 groups during the Panadol investigation. The common theme that emerged was that students were not satisfied with the apparent trend indicated by their initial results or that the results did not fit their pre-conceived idea of what the trend should be. The approach ranged from repeating selected measurements that appeared to not

fit the trend (Case study 4) to discarding all trials and repeating them (Case study 6). The fluctuation in data was often explained by the groups as mistakes that they had made (eg Case study 4) rather than being due to random error. This highlights the students' lack of understanding of the effect of random error and the lack of skills in investigating to reduce this error (Allie *et al*, 1998; Valeras, 1997).

Skills of Measurement

The manipulative skills of measuring liquid volume, like having the container rest vertically on a bench, reading the volume at eye level and reading from the bottom of the meniscus, are fundamental skills that are taught very early in most science courses. Many science programs would not provide opportunities for further practise of these skills with feedback to correct the students' techniques and improve their skills (Hodson, 1993). Continuum 7 reveals a range of performance in this skill area and indicates no real progressional trend of skill development with age.

Using equipment consistently is one of the less complex skills of level 3 of the *Working Scientifically* strand of the Western Australian Outcomes and Standards Framework (Education Department of Western Australia, 1998), and therefore should be mastered early in junior high school. The results show that many groups did not display this consistency both from student to student and for individual students. It was not uncommon to observe a student reading the volume at eye level, for example, on one occasion and then not reading at eye level for the next measurement. There

were even occasions when students were reminded by the other group members of the correct technique, yet still persisted with the inconsistent behaviour themselves.

Another disturbing and significant feature of the groups' performances with the skills of measurement was that despite not reading volumes correctly, when asked in the debriefing interview how they had ensured that their volume measurements were accurate, they had quoted reading the bottom of the meniscus, reading at eye level and having the container vertical on the bench. Whether a student was reading the bottom of the meniscus or not was difficult to determine, however students did say that they had read the bottom of the meniscus when they had not read the volume at eye level. The students' willingness to quote what they have been taught as the correct procedure, even when they have clearly not demonstrated that procedure indicates that whilst they are aware of what they should be doing they do not use the correct technique. This may indicate that they do not really believe that these procedures will enhance the accuracy of their measurements to any great extent.

Further evidence of the above claim can be gleaned from examining the groups' reasons for their equipment choice in conjunction with the demonstrated measurement skills for that group. This reveals that several of the groups that had indicated that their choice of equipment was to enhance the accuracy of their measurements demonstrated very poor measurement skills, despite the fact that these groups claimed that they had engaged in the correct technique for measuring liquid volume (eg Group 12-biol). The broad implication is that the practical measurement

skills that they have obviously been taught, as they are aware of them, were not applied to these investigation tasks.

Understanding Uncertainty

Perhaps the single most important aspect of conducting investigations is the understanding that a student has about the uncertainty associated with the data that is collected (Lubben & Millar, 1996). Knowledge of this will guide planning for an investigation, ensure care when making measurements and provide a sound basis on which to draw appropriately qualified conclusions and make recommendations (Hackling & Garnett, 1995). The three continua dealing with the students' understanding of uncertainty (Continua 8, 9 & 10) show some interesting trends within themselves but also some pertinent points can be seen when they are compared and contrasted.

Continuum 8 shows a very strong trend in that younger students are very confident in their recommendations and this confidence is diminished in the older students. Recommendations like "Use a 150mL sized cup" (Group 8-2) and "Buy Huggies" (Group 8-3) were typical of the younger students, whilst the older students made recommendations like "Between 60mL and 100mL shows the quickest time for dissolving" (Group 12-chem) and "Name brand nappies appear to absorb more water than home brand nappies" (Group 12-biol). It is interesting that even though some groups were very reluctant to admit any confidence in their results they were still willing to stand by the recommendations they had made. Traditional practical work

demands of the students that a conclusion or summation for the activity be generated, and despite not being confident with what they had done felt obligated to present some finished form at the end of the investigation (Fairbrother & Hackling, 1997). The students were certainly not prepared to concede that they were unable to make a recommendation based on the uncertainty associated with their data.

Continuum 10 maps the reasons students gave for having the level of confidence in their recommendations that they had previously expressed. It is interesting to note that none of the groups performed any kind of statistical analysis, such as averaging their data or comparing the range of data for different tests, and simply relied on the graphs that they constructed or even just the numbers themselves. The one group that had indicated that their confidence in their recommendations was supported by the trends in the data had drawn a graph of the data for the Panadol investigation but not for the nappy investigation and performed only single trials for both investigations (Case study 5).

Of the two groups that had performed replicate trials neither graphed the data or even averaged the trials and both groups indicated that the numerical differences in their data were enough to be conclusive. The majority of the groups indicated that they were confident in their recommendations because their data was conclusive, making such statements as "It's obvious" and "Our test proves it".

With the exception of the two groups that could not identify any sources of error in the investigations that they had undertaken, most groups could identify at least some source of error. The difference between younger and older groups was in the type of response given to questions about possible sources of error, the younger students were willing to rattle off such phrases as 'human error', 'equipment error' and 'measurement error' but made no attempt to explain these points in any detail. In contrast the older students, one Year 10 group and the Year 12 groups, were using the same language but were more prepared to explain in detail what they meant by each type of error. When asked what factors would have affected their data they gave responses like "Human error, it was hard to judge when the water was absorbed by the nappy" (Case study 3) "Measurement error because we didn't really measure the volumes exactly" (Case study 6). It seems likely then that the younger students have rote learned responses to the question of errors without really understanding them, whilst the older students, with more experience, are able to explain or at least give examples of the errors. This supports the claim by Allie *et al* (1998) that lack of understanding of the types of errors involved in measurement leads to a haphazard use of terminology, such as 'equipment error', 'human error' and 'measurement error'.

When the three continua dealing with understanding uncertainty are considered together some interesting points arise. As would be expected, the two groups that could identify no errors in their investigation were also very confident in the recommendations that they had made and also gave very simple responses to questions about the reasons for their confidence. One interesting group, however, is

group Year 8-1. This group had made recommendations on the basis that their trial data showed obvious conclusions, and had stated that they were very confident with the recommendations that they had made. However they were able to identify several sources of error in the investigations that they conducted. This indicates that these students have no belief that the errors actually make any real or substantial difference to the data collected.

Another group that stands out when analysing the three continua is group Year 12-biol. This group had made recommendations on the basis that they were “obvious” from their data, yet were able to identify many errors including a sampling error. Upon examining their data, their recommendation of best volume of water for dissolving Panadol was 175mL, taking 136s to dissolve the tablet. However volumes of 100mL, 125mL, 150mL, 200mL and 250mL had dissolving times no greater than 14s above their recommended volume. Despite indicating several sources of error and stating that they were prepared to accept that their recommendation was sound, they ignored the narrow range of time for their recommended volume to other measured volumes above and below that recommended. Therefore they are prepared to make this definitive judgment whilst being aware of many factors that may have influenced their results. Again it may be that they simply do not believe that errors can make a significant difference to measurements, although they did concede that they were not confident in their results but prepared to accept the validity of their recommendation.

In consideration of the four areas examined in this chapter, there was evidence that the students showed a range of skills and understanding both within an age group and across the age groups. Similarly the students demonstrate a range of performance on the different skills involved in investigation work. The skills that are most prevalent in traditional science courses, such as skills of measurement, were demonstrated with a relatively high degree of competence, whilst the skills that the traditional science course do not address, such as planning investigations, were not demonstrated at the same level of competence.

Developmental Trends

To conclude the discussion of results it is appropriate to compare the developmental trends described by the continua with the levels of student outcomes in the Outcomes and Standards Framework (Education Department of Western Australia, 1998).

For planning investigations the students mapped on the continua display a range of achievement in the outcomes from Year 8 to Year 12, and also a range within age groups. Groups 8-1 and 10-2 correspond to level 1 for planning whilst the remaining groups demonstrate skills associated with levels 2 and 3. Only groups 12-chem and 12-phys demonstrated skills associated with level 4 and some of the skills of level 5.

The levels for *Conducting* are a little more difficult to compare to the continua as the continua contain much more detail than do the outcome descriptors, and a level of performance on the outcome statements is dependent on which of the continua are used for comparison. When considering Continuum 4 two of the Year 8 groups and all of the Year 10 groups are restricted to level 2 because of the reference to discrete data in the descriptor for that level. However when comparing Continuum 6 to the outcome statements only two groups, 12-chem and 12-phys, demonstrate skills beyond level 3 on the basis of replicated trials and preliminary trials.

When considering the levels for *Evaluating* it might be pertinent to work backwards to identify the skills in these levels in the groups mapped on the continua. As none of the groups made any suggestions for improvements to their investigations none are demonstrating level 4 skills. Level three refers to students identifying difficulties with their investigations and this can be seen in the Year 12 groups as well as 10-2 when compared with Continuum 9. Therefore by elimination the remaining groups are demonstrating skills no higher than level 2 for this outcome.

Possibly the most pertinent aspect of this comparison is the implied linear progression in the student outcome statements, which can be identified in some of the continua but some of the continua do not support this notion. For example group 10-2 had discussed alternative procedures and planned for accurate measurements (level 6) but gave no consideration in their planning of interfering variables (level 3). Therefore for the sub-strand of *planning investigations* these students are displaying behaviour

at two vastly different levels of the one sub-strand. Similarly group 10-1 demonstrated great care to avoid parallax error (level 6) but made no replication of trials (level 3), thus for one aspect of the sub-strand *conducting investigations* these students could be placed at level 6 however if another aspect of the same sub-strand is considered they are performing at level 3. Therefore it is clear that the linear progression implied by the outcome statements cannot be applied across the entirety of the one investigation and is dependent upon which particular behaviour a teacher will focus on.

CHAPTER 6: CONCLUSION

Limitations

Due to the nature of this study it has some limitations that must be considered when drawing conclusions from the case studies. The study is limited in its sample size in that it used only three representative groups from each of the three year levels, Years 8, 10 and 12. The study is further limited by the fact that the students were from a specific demographic group being private school students from a middle class socioeconomic area. The sample limitations would affect the extent to which findings could be generalised, but that was not the purpose of the study. The purpose of this case study approach was to obtain rich descriptions of the situations and contexts in which students make their decisions, and to then identify the reasoning behind those decisions.

The study was cross-sectional, as opposed to a longitudinal approach, and as such, does not map the progress of individual students but compares students of different age groups at the same point in time. Also, the students' performances are confined within the contexts of the specific tasks that they undertook. Students' performances must therefore be interpreted within the task and context variables of this study.

Research Findings

Research Question 1: What decisions are made about the measurement of liquid volume during the planning and data collection phases of an investigation?

Many students in this study failed to plan effectively for data collection and control of interfering variables. The Year 8 groups tended to do very little, if any, up-front planning and whilst the older groups do show a progression towards more sophisticated planning with age, the planning that they did was often disjointed and one dimensional in that they treated aspects in isolation without considering the effect of each aspect of their plan on others.

Many groups did, however, engage in planning as they were conducting their investigation, even if they had done very little up-front planning, and were prepared to make alterations to their method as problems became apparent. Again the trend was that the Year 8 groups tended to respond to specific problems in isolation whilst the older groups demonstrated an ability to consider the effect of problems on their method as a whole.

When choosing a method for the investigation, groups demonstrated an inability to consider alternative methods in terms of data collection and control of interfering variables. Only two of the Year 12 groups considered any alternatives to the method that they first planned to use. In all cases the method was decided before any other aspects were considered and data collection and control of variables were treated within the selected method. The choice of method for all students came from

prior experience, either similar tests seen on television or methods used in secondary science lessons.

When deciding on the equipment to use groups ranged from using equipment without being able to offer a reason for doing so, and in fact used different equipment for similar trials, through to choosing equipment for the degree of accuracy it afforded. There was an age progression for decisions about equipment, in that the older students tended to consider the degree of accuracy afforded by equipment whereas younger students tended to have no reason for their choice or chose equipment for the convenience of its shape or graduations. No students considered their choice of equipment in terms of the appropriate degree of accuracy required for their investigation.

Planning to control interfering variables ranged from nonexistent or, at best, haphazard, through to controlling a number of the interfering variables associated with their investigations. In general, students tended to consider variables that were the result of learned responses to traditional practical sessions without considering whether the effect of the variables was important or not. Additionally some groups appeared to plan the control of some variables whilst ignoring others that may have had a much greater effect on their data. In the debriefing interviews several students indicated that they did not believe that the variables, which they had not controlled, would significantly affect their data.

When students were collecting the data for their investigations several issues arose. Firstly, all age groups displayed a haphazard approach when it came to dealing with mistakes in that they were reluctant to disregard and repeat a trial and happy to continue even when mistakes were identified. This is most likely a symptom of traditional practical work and indicates that students are adept at fitting their data to a preconceived outcome. Secondly, there was a tendency for groups to reduce a continuous variable to a discrete variable, although this tendency was not as strong in the older students as in the younger students.

Another significant aspect of the planning and data collection was the students' reluctance, or lack of awareness of the need, to perform replication of trials. This is the most obvious indicator of the students' lack of understanding of the effect of experimental error, and even the groups that did replicate their trials struggled to articulate their reasons for performing replicate trials. This may be attributed to the students' background of traditional practical activities where time is often limited and their reliance on completing an activity and arriving at an expected conclusion.

In general all of the students demonstrated that they lacked a schema to guide the systematic planning of their investigations, and many students seemed not to recognise the importance of thorough planning. They also displayed an inability to recognise the connection between different aspects of their procedures and the experimental error inherent in those procedures.

In summary the developmental trends for planning were that Year 8 students engaged in little or no up-front planning whilst older students were prepared to spend more time on up-front planning. The younger groups (Years 8 and 10) did not consider alternatives to their design whilst the Year 12 groups demonstrated a willingness to consider alternatives. Finally the year 12 groups and some of the Year 10 groups were more adept at choosing equipment for accuracy whilst the Year 8 groups tended to choose equipment for its convenience of shape or graduations.

Research Question 2: To what extent have students attained competence in measuring liquid volume, and is there any evidence of progression in skill development from Years 8 to 12?

The study reveals no real progression in skills of measuring liquid volume from younger students to older students. Both good and poor skills in measuring liquid volume could be observed in all age groups. Whilst some groups demonstrated good measurement skills, some aspects of the students' measurement skills are of concern. First, all year groups displayed inconsistent behaviour when measuring liquid volume, even the students who demonstrated good technique were prone to lapses. Second, many students were prepared to cite the correct technique as evidence that they had ensured their volume measurements were correct, even though they hadn't demonstrated this behaviour. This is an indicator that the students are aware of the correct procedure as a result of the science course they had undertaken, but had not used the correct technique, perhaps because they did not believe it would make much difference to the accuracy of their results. This is of particular concern as these same

students were happy to cite their measurement of volume as a source of error, indicating that this is a learned response, yet a response they do not fully understand.

Research Question 3: To what extent do students understand the uncertainty associated with the volume data they have collected, and does this affect the confidence they have in their conclusions based on these data and is there any evidence of progression in understanding from Years 8 to 12?

Students did show a trend of progression in the level of confidence that they had with their conclusions/recommendations. The Year 8 students were totally confident in their recommendations, whilst the Year 10 and Year 12 groups were progressively less confident in their recommendations. Subsidiary to this, despite being aware of errors associated with the data, and having progressively less confidence in the recommendations (with age), all of the students were prepared to accept that their recommendations were reliable.

Another aspect of confidence and awareness of uncertainty is the general reluctance by all students to perform even the simplest statistical analysis of their data and to make recommendations based on their data. Even the groups that had performed replicate trials did not calculate averages and made their recommendations based on a subjective judgment about differences between brands of nappy or volumes used to dissolve Panadol.

A significant aspect of the students' consideration of uncertainty when making recommendations was that even groups that expressed a high level of confidence in their recommendations were able to cite several sources of error in their investigations.

The evidence suggests that whilst students are able to quote rote learned responses to the question of sources of errors, they appear to believe that the errors will not affect their data, or if they do, not to any significant extent. None of the groups were prepared to concede that their data were inconclusive, despite not having a high degree of confidence in their results. Students' desire to achieve finality of the investigation in the allotted time outweighs their lack of confidence in their results. This can be linked to the students having a 'get the right answer' mentality and not being prepared to concede failure.

Implications for Teaching

This study has highlighted several issues that need to be considered by teachers when planning to offer instruction in science, particularly in respect of the *Working Scientifically* outcomes in the Western Australian Curriculum Framework (Curriculum Council, 1998).

Students in this study have demonstrated poor planning skills for investigation work and a lack of a planning schema that may be applied generally to investigations. Therefore teachers need to model appropriate planning strategies and devise

investigation activities where students can develop and practise a planning schema in a number of different contexts. To this end teachers should be aware of providing a broad range of contexts within which students may conduct investigations.

The study has also shown that a group of like-aged students can demonstrate a range of skill levels in several aspects of investigating, therefore teachers need to be aware of the possible range of skill levels within a group of students. Teachers therefore need to provide open investigations that allow students to work at their level of skills and understandings, and provide formative and developmental feedback to individuals.

The study also reveals that students are unaware, or unwilling to accept, the effect of interfering variables on their data and are unaware or reluctant to use techniques to minimise the effects of experimental error on their data, such as replicating or repeating trials and calculating an average. Thus teachers need to provide opportunities for students to investigate the effect of variables and error on their data within meaningful, real-life contexts where the consequences of their conclusions and recommendations are significant.

The most disturbing evidence from the study is that the students did not use the measurement techniques they had been taught in classes to date. Even though they were aware of the techniques to reduce parallax error in measurement, they did not use those techniques. Rather than teaching these skills in isolation and free of context

they should be taught within a range of meaningful real-life contexts that allow students to explore and investigate the effect of these techniques on the accuracy of their measurements.

When considering the use of the student outcome statements within the Outcomes and Standards Framework (Curriculum Council, 1998) for assessing performance in investigation work teachers need to be wary of the implied linear progression within these outcome statements, as it has been shown that the skills of higher levels do not necessarily presuppose achievement of skills in the preceding levels.

Implications for Further Research

Due to the small sample in this study it is difficult to generalise about students of similar ages in other schools. Therefore there is a need for a larger study in order to establish whether these findings are typical or atypical of the broader population.

The study is unable to map the progress of individual students as they progress through secondary schooling. Therefore a longitudinal study is needed to adequately report progression of investigation skills with age. This would be of benefit not only to validate the findings of this study but to give a more detailed description of the progression of skills with age and identify some of the factors that influence the development of these skills.

As performance in investigation skills reported in this study is confined to the tasks and contexts in which they were undertaken it would be of benefit to compare students' performances on a wider range of tasks and identify aspects of performance that are generic across tasks, and aspects that are strongly influenced by task and context variables.

As the students that participated in this study had a background of traditional recipe-style secondary science laboratory work two further areas of research emerge. This study needs to be compared with similar studies of students who have had exposure to investigation work in their science lessons, and, perhaps more importantly, an intervention study needs to be conducted to see if it is possible to devise learning experiences that address the deficiencies in students' performances highlighted by this research.

In Conclusion

This study has highlighted that students lack key investigation skills even after completing five years of secondary science education. It is evident that skills identified as being important in various curriculum documents are not yet being developed within schools. Therefore there is an urgent need to reform the implemented science curriculum to address these failings, and provide teachers with professional development experiences that will assist them to implement investigation activities and other strategies that are effective in facilitating students' development of these skills.

References

- Allie, S., Buffler, A., Kaunda, L., Campbell, B. & Lubben, F. (1998). First-year physics students' perceptions of the quality of experimental measurement. *International Journal of Science Education*, 20(4), 447-459.
- Australian Education Council. (1994). *Science: A curriculum profile for Australian Schools*. Victoria: Curriculum Corporation.
- Brown, M., Blondel, E., Simon, S. & Black, P. (1995). Progression in measurement. *Research Papers in Education*, 10(2), 143-170.
- Chi, M. T. H., Feltovich, P. J. & Glaser, R. (1981). Categorisation and representation of physics problems by experts and novices. *Cognitive Science*, 5, 121-152.
- Cohen, L. & Manion, L. (1994). *Research methods in education*. London: Routledge.
- Coles, M. & Gott, R. (1993). Teaching scientific investigation. *Education in Science*, 30, 8-11.
- Collins, A., Brown, J.S. & Newman, S.E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing and mathematics. In L. B. Resnick (Ed.), *Cognition and instruction: Issues and agendas*. Hillsdale, NJ: Erlbaum.
- Curriculum Council. (1998). *Curriculum framework for kindergarten to Year 12 education in Western Australia*. Osborne Park: Curriculum Council.
- Curriculum Council. (1999). *Syllabus manual: Year 11 and Year 12 Subjects*. Osborne Park: Curriculum Council.

Department of Education and Science. (1983). *Science in schools: Age 13 (Report No. 2)*. London: HMSO.

Department of Education and Science. (1985). *Science at Age 15*. London: HMSO.

Duggan, S. & Gott, R. (1996a). Practical work: Its role in the understanding of evidence in science. *International Journal of Science Education*, 18(7), 791-806.

Duggan, S. & Gott, R. (1996b). Scientific evidence: The new emphasis in the practical science curriculum in England and Wales. *The Curriculum Journal*, 7(1), 17-33.

Education Department of Western Australia (1994). *Profiles of student achievement: Student performance in science in Western Australian government schools*. Perth: Education Department of Western Australia.

Erickson, G., Bartley, T., Carlisle, R., Meyer, K. & Stavy, R. (1992). *The British Columbia assessment of science, part II: Student performance component*. Victoria, B.C.: Queen's Printer.

Fairbrother, R. & Hackling, M.W. (1997). Is this the right answer? *International Journal of Science Education*, 19(8), 887-894.

Garnett, P. (1998). *Science laboratory investigations in Year 9: Instruction and assessment*. Unpublished doctoral thesis, Edith Cowan University, Perth, Western Australia.

Garnett, P.J., Garnett, P.J. & Hackling, M.W. (1995) Refocusing the chemistry lab: A case for laboratory-based investigations. *Australian Science Teachers Journal*, 41(2), 26-32.

Gott, R., Duggan, S., Millar, R. & Lubben, F. (1995). Progression in investigative work in science. In P. Murphy, M. Sellinger, J. Bourne, & M. Briggs (Eds.) *Subject learning in the primary curriculum* (pp. 184-196). London: Routledge.

Hackling, M.W. & Fairbrother, R. (1996). Helping students to do open investigations in science. *Australian Science Teachers Journal*, 42(4), 26-33.

Hackling, M.W. & Garnett, P.J. (1995). The development of expertise in science investigation skills. *Australian Science Teachers Journal*, 41(4), 80-86.

Hackling, M.W., Garnett, P.J., Fairbrother, R. & Tunks, A. (1997, July). Progression in learning competencies associated with measurement and uncertainty of data. Paper presented at the 28th annual conference of the Australasian Science Education Research Association, Adelaide.

Hodson, D. (1988, August). Theoretically unsound and pedagogically useless: A critical look at practical work in school science. Paper presented at SCICON 88 conference of the New Zealand Science Teachers Association, Auckland, New Zealand.

Hodson, D. (1990). A critical look at practical work in school science. *School Science Review*, 71(256), 33-40.

Hodson, D. (1992). Redefining and reorienting practical work in school science. *School Science Review*, 73(264), 65-78.

Hodson, D. (1993). Rethinking old ways: towards a more critical approach to practical work in school science. *Studies in Science Education*, 22, 85-142.

Jones, A., Simon, S., Fairbrother, R., Watson, R. & Black, P.J. (1992). *Development of open work in school science*. Hatfield: Association for Science Education.

Larkin, J. H. (1979). Processing information for effective problem solving. *Engineering Education*, 70, 285-288.

Lucas, A.M. & Tobin, K. (1987). Problems with "Control of Variables" as a process skill. *Science Education*, 71(5), 685-690.

Lubben, F. & Millar, R. (1996). Children's ideas about the reliability of experimental data. *International Journal of Science Education*, 18(8), 955-968.

Mayer, E. (1992). *Putting general education to work: The key competencies report*. Canberra: Australian Education Council.

McGaw, B. (1986). Turning novices into experts. *Panorama*, 7, 1, 24-26.

National Academy of Sciences - National Research Council. (1996). *National science education standards*. National Academy Press: Washington D.C.

Osborne, J. (1993). Alternatives to practical work. *School science Review*, 75(271), 117-123.

Osborne, R.J., Bell, B.F. & Gilbert, J.K. (1983). Science teaching and children's views of the world. *European Journal of Science Education*, 5(1), 1-14.

Rigano, D.L. & Ritchie, S.M. (1995). Student disclosures of fraudulent practice in school laboratories. *Research in Science Education*, 25(4), 353-363.

Roth, W-M. & Roychoudhury, A. (1993). The development of science process skills in authentic contexts. *Journal of Research in Science Teaching*, 30, 127-152.

Science Functional Expert Group of the OECD (1998). *Framework for assessing scientific literacy*. OECD PISA national project.

Sere, M-G., Journeaux, R. & Larcher, C. (1993). Learning the statistical analysis of measurement error. *International Journal of Science Education*, 15(4), 427-438.

Staer, H., Goodrum, D. & Hackling, M. (1998). High school laboratory work in Western Australia: Openness to inquiry. *Research in Science Education*, 28(2), 219-228.

Stake, R. (1988). Case study methods in educational research: Seeking sweet water. In R.M. Jaeger (Ed.), *Complementary methods for research in education*, Washington: American Educational Research Association.

Stake, R. (1994). Case studies. In N. Denzin & G. Leder (Eds.), *Handbook of qualitative research*. London: Sage.

Tamir, P. (1993). A focus on student assessment. *Journal of Research in Science Teaching*, 30(6), 535-536.

Tamir, P. & Amir, R. (1987). Inter-relationship among laboratory process skills in biology. *Journal of Research in Science Teaching*, 24(2), 137-143.

Tamir, P. & Lunetta, V.N. (1981). Inquiry related tasks in high school laboratory handbooks. *Science Education*, 65, 477-484.

Toh, K-A, & Woolnough, B.E. (1990). Assessing, through reporting, the outcomes of scientific investigations. *Educational Research*, 32(1), 59-65.

Valeras, M. (1997). Third and fourth graders' conceptions of repeated trials and best representatives in science experiments. *Journal of Research in Science Teaching*, 34(9), 853-872.

Watson, R. & Fairbrother, R. (1993). Open-ended work in science (OPENS) project: Managing investigations in the laboratory. *School Science Review*, 75(271), 31-38.

Woolnough, B. & Allsop, T. (1985). *Practical work in science*. Cambridge: Cambridge University Press.

Appendix 1: Opening statement

Thank you for giving up your time this afternoon to assist me in my research. I am interested in how students plan and carry out investigations to solve problems, and I will be asking you as a group to carry out two investigations.

Please read the task statements at the top of each sheet which will give you the problem you are to investigate, feel free to ask questions to clarify the statement and to ask for any equipment that you wish to use that is not already here.

Appendix 2: Equipment List

3 brands of nappies for same age child

1 packet of soluble Panadol

3 × 9L buckets (same colour)

1 × 1L measuring cylinder

2 × 500mL measuring cylinders

2 × 100mL measuring cylinders

2 × 50mL measuring cylinders

2 × 20mL measuring cylinders

2 × 10mL measuring cylinders

2 × 1L beakers

4 × 500mL beakers

4 × 250mL beakers

4 × 100mL beakers

4 × 50mL beakers

6 stopwatches

2 funnels

4 pipettes

1 roll of paper towel

Appendix 3: Planning and Report Sheet

Which is the Best Brand of Nappy?

Task: Conduct an investigation to find out which brand of nappy will hold the greatest volume of water/urine. Collect sufficient data that will allow you to confidently advise a friend which brand they should buy as the most absorbent.

What are you going to investigate?

What is the plan for your experiment?

- Outline your general approach and the equipment you will use.

- Say how you will make sure your tests are fair and/or how variables will be controlled.

- say what measurements you will take.

Record your results here.

Will it be helpful to present your results as a graph?

What do your results tell you?

Are there any patterns or trends in your results?

What recommendation can you make to your friend based on your results?

Can you justify your recommendation?

Appendix 4: Interview Questions

General

- Q1.** How confident are you with your recommendations?
- Q2.** What are the main sources of error in your experiments?
- How important are these to the data you collected?

Task 1

- Q1.** Why did you decide on that method to investigate how much water each nappy would hold?
- Q2.** (no replication) I noticed that you measured the volume of water absorbed by one nappy of each type. If you repeated the experiment on three nappies of the same type would you get the same result? Why/why not?
- (with replication) I noticed that you measured the amount of water absorbed by 2/3 nappies of each type, why?
- (If more accurate/reliable) Why is that?
- Q3.** Why did you decide to use that particular equipment to measure the amount of water each nappy absorbed?
- Q4.** How did you make sure that your measurements of volume were accurate?

Task 2

- Q1.** Why did you decide on that method to investigate the effect of water volume on how long it took the panadol to dissolve?
- Q2.** (no replication) I noticed that you measured the time to dissolve once for each different volume. If you repeated the experiment three times for the same volume of water would you get the same result? Why/why not?
- (replication) I noticed that you measured the time for dissolving 2/3 times for each volume of water, why?
- (If more accurate/reliable) Why is that?
- Q3.** Why did you decide to use that particular equipment to measure the volumes of water that the panadol was dissolved in?
- Q4.** How did you make sure your measurements of volume were accurate?